PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/24, 15/00, 15/85, A61K 38/20, C07K 16/00, C12P 21/02, G01N 33/53, 33/68, C12Q 1/68

A1

(11) International Publication Number: WO 99/32632

(43) International Publication Date: 1 July 1999 (01.07.99)

(21) International Application Number: PCT/US98/27068

(22) International Filing Date: 18 December 1998 (18.12.98)

(30) Priority Data:
08/994,890 19 December 1997 (19.12.97) US

(71) Applicant: MILLENNIUM BIOTHERAPEUTICS, INC. [US/US]; 620 Memorial Drive, Cambridge, MA 02139 (US).

(72) Inventor: HOLTZMAN, Douglas, A.; 76 Fairmont Street #1, Cambridge, MA 02139 (US).

(74) Agents: MANDRAGOURAS, Amy, E.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US) et al.

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NOVEL EMBRYO–DERIVED INTERLEUKIN RELATED FACTOR MOLECULES AND USES THEREFOR

(57) Abstract

Novel EDIRF polypeptides, proteins, and nucleic acid molecules are disclosed. In addition to isolated, full–length EDIRF, proteins, the invention further provides isolated EDIRF fusion proteins, antigenic peptides and anti–EDIRF antibodies. The invention also provides EDIRF nucleic acid molecules, recombinant expression vectors containing a nucleic acid molecule of the invention, host cells into which the expression vectors have been introduced and non–human transgenic animals in which an EDIRF gene has been introduced or disrupted. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NOVEL EMBRYO-DERIVED INTERLEUKIN RELATED FACTOR MOLECULES AND USES THEREFOR

Background of the Invention

Cytokines are small peptide molecules produced by a variety of cells that mediate a wide range of biological activities. Arai, K.-I. *et al.* (1990) *Annu. Rev. Biochem.* 59:783 and Paul, W.E. and R.A. Seder (1994) *Cell* 76:241. Through a complex network, cytokines regulate functions including cellular growth, inflammation, immunity, differentiation and repair. Mosmann, T.R. (1991) *Curr. Opin. Immunol.* 3:311. One family of cytokines, termed interleukin-17 (IL-17), has been identified which plays a role in regulation of immune function. IL-17 family members identified to date include human IL-17 (see Yao *et al.* (1995) *J. Immunol.* 155:5483-5486 and Fossiez *et al.* (1996) *J. Exp. Med.* 183:2593-2603), murine IL-17 (see Yao *et al.* (1996) *Gene* 168:223-225), alternatively known as cytotoxic T lymphocyte associated antigen-8 (CTLA-8), rat IL-17, also known as CTLA-8 (see Rouvier *et al.* (1993) *J. Immunol.* 150:5445-5456), and a viral IL-17 known as Herpesvirus Saimiri gene 13 (HSV13) (see Albrecht *et al.* (1992) *J. Virol.* 66(8):5047-5058 and Yao *et al.* (1995) *Immunity* 3:811-821.) The IL-17 cytokines are known to induce stromal cells to secrete cytokines involved in inflammatory and hematopoietic processes. For example, IL-17 has been shown to induce secretion of IL-6, IL-8, PGE2, and G-CSF from primary cultures of synovial fibroblasts as well as inducing IL-6 secretion in six additional stromal cell lines tested. Fossiez *et al.* (1996) *J. Exp. Med.* 183:2593-2603. The induction of IL-6 secretion by fibroblasts has been correlated with both activation of NF-κB and induction of surface expression of intracellular adhesion molecule-1 (ICAM-1) as well as with a mitogenic response in T cells. Yao *et al.* (1996) *Gene* 168:223-225 and Yao *et al.* (1995) *Immunity* 3:811-821. Furthermore, IL-17 has been implicated indirectly in the hematopoiesis of neutrophils. Fossiez *et al.* (1996) *J. Exp. Med.* 183:2593-2603.

Given the important role of cytokines, such as the IL-17 cytokines, in regulation of the cellular immune response, inflammation and hematopoietic processes, there exists a need for the identification of novel IL-17-like molecules as well as for modulators of such molecules for use in regulating a variety of cellular responses.

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel IL-17-like molecules, referred to herein as Embryo-Derived Interleukin Related Factor (EDIRF) nucleic acid and protein molecules. The EDIRF molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes.

Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding EDIRF proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of EDIRF-encoding nucleic acids.

In one embodiment, an EDIRF nucleic acid molecule is 60% homologous to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof. In another embodiment, an EDIRF nucleic acid molecule is 60% homologous to the nucleotide sequence shown in SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, or a complement thereof. In a preferred embodiment, an isolated EDIRF nucleic acid molecule encodes the amino acid sequence of human or mouse EDIRF.

In a preferred embodiment, an isolated EDIRF nucleic acid molecule has the nucleotide sequence of SEQ ID NO:7 or a complement thereof. In another embodiment, an EDIRF nucleic acid molecule further comprises nucleotides 1-68 of SEQ ID NO:1. In yet another preferred embodiment, an EDIRF nucleic acid molecule further comprises nucleotides 609-738 of SEQ ID NO:1.

In another preferred embodiment of the invention, an isolated EDIRF nucleic acid molecule has the nucleotide sequence of SEQ ID NO:8 or a complement thereof. In another embodiment, an EDIRF nucleic acid molecule further comprises nucleotides 1-73 of SEQ ID NO:3. In yet another preferred embodiment, an EDIRF nucleic acid molecule further comprises nucleotides 613-756 of SEQ ID NO:3.

In yet another preferred embodiment of the invention, an isolated EDIRF nucleic acid molecule has the nucleotide sequence of SEQ ID NO:9 or a complement thereof. In another embodiment, an EDIRF nucleic acid molecule further comprises nucleotides 1-128 of SEQ ID NO:5. In yet another preferred embodiment, an EDIRF nucleic acid molecule further comprises nucleotides 912-1259 of SEQ ID NO:5.

In yet another preferred embodiment, an isolated EDIRF nucleic acid molecule has the nucleotide sequence shown SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof.

In another embodiment, an EDIRF nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, or SEQ ID NO:6. In yet another embodiment, an EDIRF nucleic acid molecule includes a nucleotide sequence

encoding a protein having an amino acid sequence at least 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6. In a preferred embodiment, an EDIRF nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO: 2, SEQ ID NO:4 or SEQ ID NO:6.

In another embodiment, an isolated nucleic acid molecule of the present invention encodes an EDIRF protein which includes a signal sequence, an N-terminal unique domain, a C-terminal IL-17-like domain, and is secreted. In another embodiment, the EDIRF nucleic acid molecule encodes an EDIRF protein and is a naturally occurring nucleotide sequence. In yet another embodiment, an isolated nucleic acid molecule of the present invention encodes an EDIRF protein and comprises a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

Another embodiment of the invention features EDIRF nucleic acid molecules which specifically detect EDIRF nucleic acid molecules relative to nucleic acid molecules encoding non-EDIRF proteins. For example, in one embodiment, an EDIRF nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising nucleotides 520-738 of SEQ ID NO:1. In another exemplary embodiment, an EDIRF nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising nucleotides 525-756 of SEQ ID NO:3. In another embodiment, an EDIRF nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule comprising nucleotides 1-350 or to nucleotides 1025-1259 of SEQ ID NO:5. In another embodiment, the EDIRF nucleic acid molecule is at least 500 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof.

Another embodiment the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of an EDIRF nucleic acid.

Another aspect of the invention provides a vector comprising an EDIRF nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment the invention provides a host cell containing a vector of the invention. The invention also provides a method for producing EDIRF protein by

culturing in a suitable medium, a host cell of the invention containing a recombinant expression vector such that EDIRF protein is produced.

Another aspect of this invention features isolated or recombinant EDIRF proteins and polypeptides. In one embodiment, an isolated EDIRF protein has a signal sequence, an N-terminal unique domain, a C-terminal IL-17-like domain, and is secreted. In another embodiment, an isolated EDIRF protein has an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6. In a preferred embodiment, an EDIRF protein has an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6. In another embodiment, an EDIRF protein has the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6. In yet another embodiment, an isolated EDIRF protein is a cell surface-associated protein.

Another embodiment of the invention features an isolated EDIRF protein which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 60% homologous to a nucleotide sequence of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof. Another embodiment of the invention features an isolated EDIRF protein which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 60% homologous to a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof. This invention also features an isolated EDIRF protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, or a complement thereof. The invention further features an isolated EDIRF protein which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof.

The EDIRF proteins of the present invention, or biologically active portions thereof, can be operatively linked to a non-EDIRF polypeptide to form EDIRF fusion proteins. The invention further features antibodies that specifically bind EDIRF

proteins, such as monoclonal or polyclonal antibodies. In addition, the EDIRF proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting EDIRF expression in a biological sample by contacting the biological sample with an agent capable of detecting an EDIRF nucleic acid molecule, protein or polypeptide such that the presence of EDIRF nucleic acid molecule, protein or polypeptide is detected in the biological sample.

In another aspect, the present invention provides a method for detecting the presence of EDIRF activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of EDIRF activity such that the presence of EDIRF activity is detected in the biological sample.

In another aspect, the invention provides a method for modulating EDIRF activity comprising contacting the cell with an agent that modulates EDIRF activity such that EDIRF activity in the cell is modulated. In one embodiment, the agent inhibits EDIRF activity. In another embodiment, the agent stimulates EDIRF activity. In one embodiment, the agent is an antibody that specifically binds to EDIRF protein. In another embodiment, the agent modulates expression of EDIRF by modulating transcription of an EDIRF gene or translation of an EDIRF mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of the EDIRF mRNA or the EDIRF gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant EDIRF protein or nucleic acid expression or activity by administering an agent which is an EDIRF modulator to the subject. In one embodiment, the EDIRF modulator is an EDIRF protein. In another embodiment the EDIRF modulator is an EDIRF nucleic acid molecule. In yet another embodiment, the EDIRF modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant EDIRF protein or nucleic acid expression is an immune disorder, a hematopoietic disorder, a differentiative or developmental disorder, or a proinflammatory disorder.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding an EDIRF protein; (ii) mis-regulation of said gene; and (iii) aberrant post-translational modification of an EDIRF protein, wherein a wild-type form of said gene encodes an protein with an EDIRF activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of an EDIRF protein, by providing a indicator

composition comprising an EDIRF protein having EDIRF activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on EDIRF activity in the indicator composition to identify a compound that modulates the activity of an EDIRF protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequences of two forms of human EDIRF. *Figure 1A* depicts the nucleotide and amino acid sequences of human EDIRF I. The nucleotide sequence corresponds to nucleic acids 1 to 738 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 180 of SEQ ID NO:2. *Figure 1B* depicts the nucleotide and amino acid sequence of human EDIRF II. The nucleotide sequence corresponds to nucleic acids 1 to 1259 of SEQ ID NO:5. The amino acid sequence corresponds to amino acids 1 to 261 of SEQ ID NO:6.

Figure 2 depicts the cDNA sequence and predicted amino acid sequence of murine EDIRF I. The nucleotide sequence corresponds to nucleic acids 1 to 756 of SEQ ID NO:3. The amino acid sequence corresponds to amino acids 1 to 180 of SEQ ID NO:4.

Figure 3 depicts an alignment of the amino acid sequences of human EDIRF I (corresponding to amino acids 1 to 180 of SEQ ID NO:2), human EDIRF II (corresponding to amino acids 1-261 of SEQ ID NO:6), murine EDIRF I (corresponding to amino acids 1 to 180 of SEQ ID NO:4), human IL-17 precursor (Swiss-Prot™ Accession No. Q16552, corresponding to amino acid residues 1-155 of SEQ ID NO:14), murine IL-17 precursor, alternatively known as cytotoxic T lymphocyte-associated antigen 8 precursor or CTLA-8 precursor (Swiss-Prot™ Accession No. Q62386, corresponding to amino acid residues 1-158 of SEQ ID NO:15), and herpesvirus Saimiri immediate early gene 13 protein precursor (Swiss-Prot™ Accession No. P24916, corresponding to amino acid residues 1-157 of SEQ ID NO:16). Cysteine residues that are conserved between EDIRF family members and IL-17 family members are indicated by astericks.)

Detailed Description of the Invention

The present invention is based on the discovery of novel molecules having homology to members of the IL-17 protein family, referred to herein as EDIRF protein and nucleic acid molecules, which comprise a family of molecules having certain

conserved structural and functional features. The nucleotide sequences of two human EDIRF nucleic acid molecules (human EDIRF I and human EDIRF II) and the amino acid sequence of the respective human EDIRF protein molecules are depicted in Figure 1. The nucleotide sequence of the murine EDIRF I nucleic acid molecule and the amino acid sequence of the murine EDIRF I protein molecule are depicted in Figure 2.

The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more protein or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin and a homologue of that protein of murine origin, as well as a second, distinct protein of human origin and a murine homologue of that protein. Members of a family may also have common functional characteristics. An exemplary family of the present invention is the IL-17 family whose members include, at least, human IL-17, murine IL-17, alternatively known as cytotoxic T lymphocyte-associated antigen 8 (CTLA-8) and herpesvirus Saimiri immediate early gene 13 protein (HSV13).

In one embodiment, an EDIRF family member is identified based on the presence of a "C-terminal IL-17-like domain" in the protein or corresponding nucleic acid molecule. As used herein, the term "C-terminal IL-17-like domain" refers to a protein domain which is at least about 70-110 amino acid residues in length, preferably at least about 80-100 amino acid residues in length, and more preferably at least about 85-90 amino acid residues in length, and has at least about 25-60%, preferably at least about 30-50%, more preferably at least about 35-45% homology, and even more preferably at least about 38-41% homology with the amino acid sequence of human IL-17, as set forth in SEQ ID NO:14.

A C-terminal IL-17-like domain further contains at least about three, more preferably at least four, and even more preferably at least five cysteine residues which are conserved between EDIRF molecules and members of the IL-17 protein family (e.g., human IL-17, murine IL-17, alternatively known as CTLA-8, and herpesvirus Saimiri immediate early gene 13 protein.) Preferably, the C-terminal IL-17-like domain of EDIRF has cysteine residues which are located in the same or similar positions as cysteine residues in a IL-17 protein family. For example, when an EDIRF protein of the invention is aligned with an IL-17 family member for purposes of comparison (see e.g., Fig. 3) preferred cysteine-rich domains of the invention are those in which cysteine residues in the amino acid sequence of EDIRF are located in the same or similar position as the cysteine residues in the IL-17 family member. As an illustrative embodiment, Fig. 3 shows cysteine residues located in the same or similar positions of the human

EDIRF protein (corresponding to SEQ ID NO:2) and human IL-17 at the following locations: amino acid number 121 of human EDIRF I and amino acid number 94 of the human IL-17 protein; amino acid number 126 of human EDIRF I and amino acid number 99 of the human IL-17 protein; amino acid number 163 of human EDIRF I and amino acid number 129 of the human IL-17 protein; amino acid number 178 of human EDIRF I and amino acid number 144 of the human IL-17 protein; and amino acid number 176 of human EDIRF I and amino acid number 146 of the human IL-17 protein.

In a preferred embodiment, a C-terminal IL-17-like domain is located in the C-terminal region of an EDIRF protein. For example, in one embodiment, an EDIRF protein contains a C-terminal IL-17-like domain containing about amino acids 95-180 of SEQ ID NO:2. In another embodiment, an EDIRF protein contains a C-terminal IL-17-like domain containing about amino acids 95-180 of SEQ ID NO:4. In yet another embodiment, an EDIRF protein contains a C-terminal IL-17-like domain containing about amino acids 176-261 of SEQ ID NO:6.

Another embodiment of the invention features EDIRF molecules which contain an N-terminal unique domain. The term "N-terminal unique domain" as used herein, refers to a protein domain of an EDIRF protein family member which includes amino acid residues N-terminal to the N-terminus of a C-terminal IL-17-like domain in the amino acid sequence of the EDIRF protein, e.g., a protein domain which includes amino acid residues from the N-terminus of the C-terminal IL-17-like domain to the N-terminal amino acid residue of the amino acid sequence of the protein. As used herein, an "N-terminal unique domain" refers to a protein domain which is at least about 70-120 amino acid residues in length, preferably at least about 80-110 amino acid residues in length, and more preferably at least about 90-100 amino acid residues in length, and has at least about 65-95%, preferably at least about 70-90%, and more preferably at least about 75-85% homology with the amino acid sequence of a second EDIRF family member. As further defined herein, a N-terminal unique domain of an EDIRF protein family member, however, is not sufficiently homologous to the amino acid sequence of a member of another protein family, such as an IL-17 protein family. For example, an N-terminal unique domain of human EDIRF I (containing about amino acids 1-94 of SEQ ID NO:2 has at least about 80% homology to the N-terminal unique domain of murine EDIRF I (containing about amino acids 1-94 of SEQ ID NO:4) but has no significant homology to the amino acid sequence of human IL-17.

As further defined herein, an N-terminal unique domain may further contain an "N-terminal signal sequence". As used herein, a "signal sequence" refers to a peptide containing about 20 amino acids which occurs at the extreme N-terminal end of secretory and integral membrane proteins and which contains large numbers of

hydrophobic amino acid residues. For example, a signal sequence contains at least about 14-28 amino acid residues, preferably about 16-26 amino acid residues, more preferably about 18-24 amino acid residues, and more preferably about 20-22 amino acid residues, and has at least about 40-70%, preferably about 50-65%, and more preferably about 55-60% hydrophobic amino acid residues (e.g., Alanine, Valine, Leucine, Isoleucine, Phenylalanine, Tyrosine, Tryptophan, or Proline). Such a "signal sequence", also referred to in the art as a "signal peptide", serves to direct a protein containing such a sequence to a lipid bilayer.

In a preferred embodiment, an EDIRF protein contains both an N-terminal unique domain and a C-terminal IL-17-like domain. In a more preferred embodiment, an EDIRF N-terminal unique domain further contains a signal sequence. In one exemplary embodiment, an EDIRF protein contains an N-terminal unique domain comprising amino acids 1-94 of SEQ ID NO:2 which further contains a signal sequence at amino acids 1-22 of SEQ ID NO:2. In another exemplary embodiment, an EDIRF protein contains an N-terminal unique domain comprising amino acids 1-94 of SEQ ID NO:4 which further contains a signal sequence at amino acids 1-22 of SEQ ID NO:4. In yet another exemplary embodiment, an EDIRF protein contains an N-terminal unique domain comprising amino acids 1-175 of SEQ ID NO:2.

Preferred EDIRF molecules of the present invention have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least about 40% homology, preferably 50% homology, more preferably 60%-70% homology across the amino acid sequences of the domains and contain at least one, preferably two, and more preferably three or four structural domains, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences that share at least 40%, preferably 50%, more preferably 60, 70, or 80% homology and share a common functional activity are defined herein as sufficiently homologous.

As used interchangeably herein an "EDIRF activity", "biological activity of EDIRF" or "functional activity of EDIRF", refers to an activity exerted by an EDIRF protein, polypeptide or nucleic acid molecule on an EDIRF responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, an

EDIRF activity is a direct activity, such as an association with a cell-surface protein (e.g., an EDIRF receptor). In another embodiment, an EDIRF activity is an indirect activity, such as the induction of synthesis of a second protein (e.g. a cellular cytokine) mediated by interaction of the EDIRF protein with a cell surface protein. In a preferred embodiment, an EDIRF activity is at least one or more of the following activities: (i) interaction of an EDIRF protein in the extracellular millieu with a non-EDIRF protein molecule on the surface of the same cell which secreted the EDIRF protein molecule; (ii) interaction of an EDIRF protein in the extracellular millieu with a non-EDIRF protein molecule on the surface of a different cell from that which secreted the EDIRF protein molecule; (iii) complex formation between an EDIRF protein and an EDIRF receptor; (iv) complex formation between an EDIRF protein and non-EDIRF receptor; and (v) interaction of an EDIRF protein with a second protein in the extracellular millieu. In yet another preferred embodiment, an EDIRF activity is at least one or more of the following activities: (i) activation of an EDIRF-dependent signal transduction pathway; (ii) modulation of secretion of a non-IL-17 cytokine; (iii) interaction of an EDIRF protein with a non-EDIRF cell surface molecule (e.g., a proteoglycan or sulfated proteoglycan); (iv) modulation of secretion of IL-17; (v) modulation of surface expression of a cellular adhesion molecule; (vi) modulation of a proinflammatory cytokine; (vii) modulation of a hematopoietic cytokine; (viii) modulation of the development or differentiation of an EDIRF-expressing cell; (ix) modulation of the development or differentiation of a non- EDIRF-expressing cell; (x) modulation of the homeostasis of an EDIRF-expressing cell; and (xi) modulation of the homeostasis of a non-EDIRF-expressing cell.

Accordingly, another embodiment of the invention features isolated EDIRF proteins and polypeptides having an EDIRF activity. Preferred EDIRF proteins have a C-terminal IL-17-like domain and an EDIRF activity. In another embodiment, the EDIRF protein has an N-terminal unique domain, a C-terminal IL-17-like domain, and an EDIRF activity. In another preferred embodiment, the EDIRF protein has at least at least a C-terminal IL-17-like domain, an N-terminal unique domain, an EDIRF activity, and an amino acid sequence sufficiently homologous to an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6.

In a particularly preferred embodiment, the EDIRF protein and nucleic acid molecules of the present invention are human EDIRF molecules. A human EDIRF cDNA molecule (human EDIRF I) was obtained from a human fetal lung cDNA library as described in Example 1. A second human EDIRF cDNA (human EDIRF II) was also obtained from the library as described in Example 1. The nucleotide sequences of the isolated human EDIRF cDNAs and the predicted amino acid sequences of the human

EDIRF proteins are shown in Figure 1 and in SEQ ID NOs:1 and 2 and SEQ ID NOs:5 and 6, respectively. In addition, the nucleotide sequence corresponding to the coding region of the human EDIRF cDNAs are represented as SEQ ID NOs:7 and 9.

EDIRF mRNA transcripts of approximately <1.0 and ~5.0kb were expressed in most human tissues tested (See Example 2).

The human EDIRF I cDNA set forth in SEQ ID NO:1, is approximately 738 nucleotides in length and encodes a protein which is approximately 180 amino acid residues in length (SEQ ID NO:2). The human EDIRF II cDNA set forth in SEQ ID NO:5, is approximately 1259 nucleotides in length and encodes a protein which is approximately 261 amino acid residues in length (SEQ ID NO:6). The human EDIRF proteins contain an N-terminal signal sequence, an N-terminal unique domain, and a C-terminal IL-17-like domain, as defined herein. An EDIRF N-terminal unique domain can be found at least, for example, from about amino acids 1-94 of SEQ ID NO:2 and, for example, from about amino acids 61-175 of SEQ ID NO:6. An EDIRF C-terminal IL-17-like domain can be found at least, for example, from about amino acids 95-180 of SEQ ID NO:2 and, for example, from about amino acids 176-261 of SEQ ID NO:6. A signal sequence can be found at least, for example, from about amino acids 1-20 of SEQ ID NO:2. The prediction of such a signal peptide can be made, for example, utilizing the computer algorithm SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

In another embodiment, the EDIRF protein and nucleic acid molecules of the present invention are murine EDIRF molecules. A murine EDIRF full-length cDNA molecule (murine EDIRF I) was identified as described in Example 1. The nucleotide sequence of the full-length murine EDIRF I cDNA and the predicted amino acid sequence of the murine EDIRF I protein are shown in Figure 2 and in SEQ ID NOs:3 and 4, respectively. In addition, the nucleotide sequence corresponding to the coding region of the murine EDIRF I cDNA is represented as SEQ ID NO:8.

EDIRF mRNA transcripts of approximately 1.0 kb were expressed at low levels in adult murine lung (See Example 2). A 1.0kb developmentally-regulated transcript was also identified in murine embryonic tissue. The expression of murine EDIRF transcripts was also detected by *in situ* analysis of 13.5-day mouse embryos.

The murine EDIRF I cDNA is approximately 756 nucleotides in length and encodes a protein which is approximately 180 amino acid residues in length (SEQ ID NO:4). The murine EDIRF I protein contains an N-terminal signal sequence, an N-terminal unique domain, and a C-terminal IL-17-like domain, as described herein. An EDIRF N-terminal unique domain can be found at least, for example, from about amino acids 1-94 of SEQ ID NO:4. An EDIRF C-terminal IL-17-like domain can be found at least, for example, from about amino acids 95-180 of SEQ ID NO:4. A signal sequence

can be found at least, for example, from about amino acids 1-22 of SEQ ID NO:4. The prediction of such a signal peptide can be made, for example, utilizing the computer algorithm SIGNALP (Henrik, et al. (1997) *Protein Engineering* 10:1-6).

An alignment of the amino acid sequences of human EDIRF I, human EDIRF II, and murine EDIRF I, as well as human IL-17 (Swiss-Prot™ Accession No. Q16552), murine IL-17, alternatively known as CTLA-8 (Swiss-Prot™ Accession No. Q62386), and herpesvirus Saimiri immediate early gene 13 protein (Swiss-Prot™ Accession No. P24916) is shown in Figure 3. (The alignment was generated using MegAlign™ sequence alignment software. The initial pairwise alignment step was performed using a Wilbur-Lipmann algorithm with a K-tuple of 2, a GAP penalty of 5, a window of 4, and diagonals saved set to = 4. The multiple alignment step was performed using the Clustal algorithm with a PAM 250 residue weight Table, a GAP penalty of 10, and a GAP length penalty of 10.)

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode EDIRF proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify EDIRF-encoding nucleic acids (e.g., EDIRF mRNA) and fragments for use as PCR primers for the amplification or mutation of EDIRF nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated EDIRF nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by

recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or portion of the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, as a hybridization probe, EDIRF nucleic acid molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to EDIRF nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to a human EDIRF cDNA. This cDNA comprises sequences encoding a human EDIRF protein (i.e., "the coding region", from nucleotides 69-608), as well as 5' untranslated sequences (nucleotides 1 to 68) and 3' untranslated sequences

(nucleotides 609 to 738). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (e.g., nucleotides 69 to 608, corresponding to nucleotides 1-540 of SEQ ID NO:7).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:3. The sequence of SEQ ID NO:3 corresponds to the murine EDIRF cDNA. This cDNA comprises sequences encoding the murine EDIRF protein (i.e., "the coding region", from nucleotides 74-613), as well as 5' untranslated sequences (nucleotides 1 to 73) and 3' untranslated sequences (nucleotides 614 to 756). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:3 (e.g., nucleotides 74 to 613, corresponding to nucleotides 1- 540 of SEQ ID NO:8).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:5. The sequence of SEQ ID NO:5 corresponds to a second human EDIRF cDNA, referred to herein as long-form human EDIRF. This cDNA comprises sequences encoding a human EDIRF protein. In one embodiment, this cDNA comprises a coding region from nucleotides 129 to 911, as well as 5' untranslated sequences from nucleotides 1 to 128 and 3' untranslated sequences from nucleotides 912 to 1259. Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:5 (e.g., nucleotides 129 to 911, corresponding to nucleotides 1- 783 of SEQ ID NO:9). The human EDIRF cDNA of SEQ ID NO:5 also contains additional in frame ATG codons at positions 264-266 and 309-311 of SEQ ID NO:5 encoding methionines at positions 46 and 61 of SEQ ID NO:6, respectively. Thus, in another embodiment, the nucleotide sequence of SEQ ID NO:5 comprises a coding region from nucleotides 264 to 911, as well as 5' untranslated sequences from nucleotides 1 to 263 and 3' untranslated sequences from nucleotides 912 to 1259. In yet another embodiment, the nucleotide sequence of SEQ ID NO:5 comprises a coding region from nucleotides 309 to 911, as well as 5' untranslated sequences from nucleotides 1 to 308 and 3' untranslated sequences from nucleotides 912 to 1259. The ATG codons of SEQ ID NO:6, at positions 1-3, 264-266, and 309-311 are in a less favorable context for translation than is the ATG codon at position 1-3 of SEQ ID NO:1, indicating that human EDIRF proteins may be translated with a lower efficiency from the long-form human EDIRF nucleotide sequence of SEQ ID NO:5 as compared to the human EDIRF nucleotide sequence of SEQ ID NO:1.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613. In another preferred embodiment, an isolated

nucleic acid molecule of the invention comprises the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614 is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 60-65%, preferably at least about 70-75%, more preferable at least about 80-85%, and even more preferably at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO:1, the nucleotide sequence shown in SEQ ID NO:3, the nucleotide sequence shown in SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO5 (or SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9), the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of EDIRF. The nucleotide sequence determined

from the cloning of the human and murine EDIRF genes allows for the generation of probes and primers designed for use in identifying and/or cloning EDIRF homologues in other cell types, e.g., from other tissues, as well as EDIRF homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 40, 50 or 75 consecutive nucleotides of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613 ,or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614 sense, or an anti-sense sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or of a naturally occurring mutant of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

Probes based on the human EDIRF nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress an EDIRF protein, such as by measuring a level of an EDIRF-encoding nucleic acid in a sample of cells from a subject e.g., detecting EDIRF mRNA levels or determining whether a genomic EDIRF gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of EDIRF" can be prepared by isolating a portion of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614 which encodes a polypeptide having an EDIRF biological activity (the biological activities of the EDIRF proteins have previously been described), expressing the encoded portion of EDIRF protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of EDIRF.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 (and portions thereof, e.g., SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9, respectively) and

from the nucleotide sequences of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613 , the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or portions thereof, due to degeneracy of the genetic code and thus encode the same EDIRF protein as that encoded by the nucleotide sequence shown in SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5 (or SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9) or by the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613 or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6.

In addition to the human and murine EDIRF nucleotide sequences shown in SEQ ID NO:1 or SEQ ID NO:5 and SEQ ID NO:3, respectively, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of EDIRF may exist within a population (e.g., the human population). Such genetic polymorphism in the EDIRF gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding an EDIRF protein, preferably a mammalian EDIRF protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the EDIRF gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in EDIRF that are the result of natural allelic variation and that do not alter the functional activity of EDIRF are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding EDIRF proteins from other species, and thus which have a nucleotide sequence which differs from the human sequences of SEQ ID NO:1 or SEQ ID NO:5, the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613 or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or the murine nucleotide sequence of SEQ ID NO:3 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the EDIRF cDNA of the invention can be isolated based on their homology to the human or murine EDIRF nucleic acids disclosed herein using the human cDNA, murine cDNA, or portions thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a viral EDIRF cDNA can be isolated based on its homology to human or murine EDIRF.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 nucleotides in length and hybridizes under stringent conditions to

the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614. In other embodiment, the nucleic acid is at least 30, 50, 100, 250 or 500 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 65%, more preferably at least about 70%, and even more preferably at least about 75% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of the EDIRF sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, thereby leading to changes in the amino acid sequence of the encoded EDIRF protein, without altering the functional ability of the EDIRF protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of EDIRF (e.g., the sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the EDIRF proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, amino acid residues that are

conserved between EDIRF protein and other proteins having IL-17-like domains are not likely to be amenable to alteration.

In addition, the amino acid sequences of human EDIRF have at least one predicted N-glycosylation site corresponding to about amino acids 75-78 of SEQ ID NO:2 or to about amino acids 156-159 of SEQ ID NO:6 which is not likely to be amenable to alteration. Likewise, the amino acid sequence of murine EDIRF has at least one predicted N-glycosylation site corresponding to about amino acids 75-78 of SEQ ID NO:4 which is not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding EDIRF proteins that contain changes in amino acid residues that are not essential for activity. Such EDIRF proteins differ in amino acid sequence from SEQ ID NO:2, SEQ ID NO:4 or to SEQ ID NO:6 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2. Preferably, the protein encoded by the nucleic acid molecule is at least about 70% homologous to SEQ ID NO:2, more preferably at least about 80% homologous to SEQ ID NO:2, even more preferably at least about 90% homologous to SEQ ID NO:2, and most preferably at least about 95% homologous to SEQ ID NO:2.

In an alternative embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:4. Preferably, the protein encoded by the nucleic acid molecule is at least about 70% homologous to SEQ ID NO:4, more preferably at least about 80% homologous to SEQ ID NO:4, even more preferably at least about 90% homologous to SEQ ID NO:4, and most preferably at least about 95% homologous to SEQ ID NO:4.

In yet another embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:6. Preferably, the protein encoded by the nucleic acid molecule is at least about 70% homologous to SEQ ID NO:6, more preferably at least about 80% homologous to SEQ ID NO:6, even more preferably at least about 90% homologous to SEQ ID NO:6, and most preferably at least about 95% homologous to SEQ ID NO:6.

An isolated nucleic acid molecule encoding an EDIRF protein homologous to the protein of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5 (or of SEQ ID

NO:7, SEQ ID NO:8 or SEQ ID NO:9), or into the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number 98613 or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in EDIRF is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an EDIRF coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for EDIRF biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant EDIRF protein can be assayed for (1) activation of an EDIRF-dependent signal transduction pathway; (2) modulation of secretion of a non-IL-17 cytokine; (3) modulation of secretion of IL-17; (4) modulation of surface expression of a cellular adhesion molecule; (5) modulation of a proinflammatory cytokine; (6) modulation of a hematopoietic cytokine; (7) modulation of the development or differentiation of an EDIRF-expressing cell; (8) modulation of the development or differentiation of a non- EDIRF-expressing cell; (9) modulation of the

homeostasis of an EDIRF-expressing cell; and (10) modulation of the homeostasis of a non-EDIRF-expressing cell.

In addition to the nucleic acid molecules encoding EDIRF proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire EDIRF coding strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding EDIRF. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the coding region of human EDIRF corresponds to SEQ ID NO:7 or to SEQ ID NO:9 the and the coding region of murine EDIRF corresponds to SEQ ID NO:8). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding EDIRF. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding EDIRF disclosed herein (e.g., SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of EDIRF mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of EDIRF mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of EDIRF mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil,

5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5- oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an EDIRF protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids.*

Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave EDIRF mRNA transcripts to thereby inhibit translation of EDIRF mRNA. A ribozyme having specificity for an EDIRF-encoding nucleic acid can be designed based upon the nucleotide sequence of an EDIRF cDNA disclosed herein (i.e., SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an EDIRF-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, EDIRF mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, EDIRF gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the EDIRF (e.g., the EDIRF promoter and/or enhancers) to form triple helical structures that prevent transcription of the EDIRF gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In preferred embodiments, the nucleic acids of EDIRF can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed

using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* PNAS 93: 14670-675.

PNAs of EDIRF can be used therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, e.g., inducing trascription or translation arrest or inhibiting replication. PNAs of EDIRF can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as 'artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup B. (1996) *supra*); or as probes or primers for DNA sequence and hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of EDIRF can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of EDIRF can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Research* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al., 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre et al., 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No. W088/09810, published December 15, 1988) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134, published April 25, 1988). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents

(See, e.g., Krol et al., 1988, BioTechniques 6:958-976) or intercalating agents. (See, e.g., Zon, 1988, Pharm. Res. 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated EDIRF Proteins and Anti-EDIRF Antibodies

One aspect of the invention pertains to isolated EDIRF proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-EDIRF antibodies. In one embodiment, native EDIRF proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, EDIRF proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an EDIRF protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the EDIRF protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of EDIRF protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of EDIRF protein having less than about 30% (by dry weight) of non-EDIRF protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-EDIRF protein, still more preferably less than about 10% of non-EDIRF protein, and most preferably less than about 5% non-EDIRF protein. When the EDIRF protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of EDIRF protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of EDIRF protein having less than about 30% (by dry weight) of chemical precursors or non-EDIRF chemicals, more preferably less than about 20% chemical precursors or non-EDIRF chemicals, still more preferably less than about 10%

chemical precursors or non-EDIRF chemicals, and most preferably less than about 5% chemical precursors or non-EDIRF chemicals.

Biologically active portions of an EDIRF protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the EDIRF protein, e.g., the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, which include less amino acids than the full length EDIRF proteins, and exhibit at least one activity of an EDIRF protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the EDIRF protein. A biologically active portion of an EDIRF protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

In one embodiment, a biologically active portion of an EDIRF protein comprises at least a C-terminal IL-17-like domain. In yet another embodiment, a biologically active portion of an EDIRF protein comprises at least a signal sequence.
In an alternative embodiment, a biologically active portion of an EDIRF protein comprises an EDIRF amino acid sequence lacking a signal sequence.

It is to be understood that a preferred biologically active portion of an EDIRF protein of the present invention may contain at least one of the above-identified structural domains. A more preferred biologically active portion of an EDIRF protein may contain at least two of the above-identified structural domains. An even more preferred biologically active portion of an EDIRF protein may contain at least three of the above-identified structural domains.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native EDIRF protein.

In a preferred embodiment, the EDIRF protein has an amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6. In other embodiments, the EDIRF protein is substantially homologous to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6 and retains the functional activity of the protein of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6 yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the EDIRF protein is a protein which comprises an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6 and retains the functional activity of the EDIRF proteins of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, respectively. Preferably, the protein is at least about 70% homologous to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, more preferably at least about 80% homologous to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6, even more preferably at least about 90% homologous to SEQ ID NO:2, SEQ ID

NO:4 or SEQ ID NO:6, and most preferably at least about 95% or more homologous to SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6.

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence and non-homologous sequences can be disregarded for comparison purposes). In one embodiment, an alignment is a global alignment, e.g., an overall sequence alignment. In another embodiment, an alignment is a local alignment. In a preferred embodiment, the length of a sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence to which it is aligned (e.g., when aligning a second sequence to the EDIRF amino acid sequence of SEQ ID NO:2, at least 54, preferably at least 72, more preferably at least 90, even more preferably at least 108, and even more preferably at least 126, 144 or 162 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, 5, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program

(version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to EDIRF nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to EDIRF protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides EDIRF chimeric or fusion proteins. As used herein, an EDIRF "chimeric protein" or "fusion protein" comprises an EDIRF polypeptide operatively linked to a non-EDIRF polypeptide. A "EDIRF polypeptide" refers to a polypeptide having an amino acid sequence corresponding to EDIRF, whereas a "non-EDIRF polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the EDIRF protein, e.g., a protein which is different from the EDIRF protein and which is derived from the same or a different organism. Within an EDIRF fusion protein the EDIRF polypeptide can correspond to all or a portion of an EDIRF protein. In a preferred embodiment, an EDIRF fusion protein comprises at least one biologically active portion of an EDIRF protein. In another preferred embodiment, an EDIRF fusion protein comprises at least two biologically active portions of an EDIRF protein. In another preferred embodiment, an EDIRF fusion protein comprises at least three biologically active portions of an EDIRF protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the EDIRF polypeptide and the non-EDIRF polypeptide are fused in-frame to each other. The non-EDIRF polypeptide can be fused to the N-terminus or C-terminus of the EDIRF polypeptide.

For example, in one embodiment, the fusion protein is a GST-EDIRF fusion protein in which the EDIRF sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant EDIRF.

In another embodiment, the fusion protein is an EDIRF protein containing a heterologous signal sequence at its N-terminus. For example, the native EDIRF signal

sequence (i.e, about amino acids 1 to 22 of SEQ ID NO:2, or about amino acids 1 to 22 of SEQ ID NO:4) can be removed and replaced with a signal sequence from another protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of EDIRF can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an EDIRF-immunoglobulin fusion protein in which the EDIRF sequences comprising primarily the EDIRF extracellular domain are fused to sequences derived from a member of the immunoglobulin protein family. Soluble derivatives have also been made of cell surface glycoproteins in the immunoglobulin gene superfamily consisting of an extracellular domain of the cell surface glycoprotein fused to an immunoglobulin constant (Fc) region (see e.g., Capon, D.J. *et al.* (1989) *Nature* 337:525-531 and Capon U.S. Patents 5,116,964 and 5,428,130 [CD4-IgG1 constructs]; Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 173:721-730 [a CD28-IgG1 construct and a B7-1-IgG1 construct]; and Linsley, P.S. *et al.* (1991) *J. Exp. Med.* 174:561-569 and U.S. Patent 5,434,131[a CTLA4-IgG1]). Such fusion proteins have proven useful for modulating receptor-ligand interactions. Soluble derivatives of cell surface proteins of the tumor necrosis factor receptor (TNFR) superfamily proteins have been made consisting of an extracellular domain of the cell surface receptor fused to an immunoglobulin constant (Fc) region (see for example Moreland *et al.* (1997) N. Engl. J. Med. 337(3):141-147; van der Poll *et al.* (1997) Blood 89(10):3727-3734; and Ammann *et al.* (1997) J. Clin. Invest. 99(7):1699-1703.) Furthermore, fusion proteins have been made using the CH2 and CH3 domains of IgG fused downstream of murine IL-17 leader sequences and upstream of murine CTLA-8 sequences and upstream of HVS13 sequences (see for example Yao *et al.* (1995) Immunity 8:811-821.)

The EDIRF-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an EDIRF protein and an EDIRF receptor on the surface of a cell, to thereby suppress EDIRF-mediated cellular function *in vivo*. The EDIRF-immunoglobulin fusion proteins can be used to affect the bioavailability of an EDIRF protein. Inhibition of the EDIRF protein/EDIRF receptor interaction may be useful therapeutically, for example, in regulation of the cellular immune response, regulation of inflammation, or regulation of hematopoiesis. Moreover, the EDIRF-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-EDIRF antibodies in a subject, to purify EDIRF receptors and in screening assays to identify molecules which inhibit the interaction of an EDIRF protein with an EDIRF receptor.

Preferably, an EDIRF chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the

different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A EDIRF-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the EDIRF protein.

The present invention also pertains to variants of the EDIRF proteins which function as either EDIRF agonists (mimetics) or as EDIRF antagonists. Variants of the EDIRF protein can be generated by mutagenesis, e.g., discrete point mutation or truncation of the EDIRF protein. An agonist of the EDIRF protein can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the EDIRF protein. An antagonist of the EDIRF protein can inhibit one or more of the activities of the naturally occurring form of the EDIRF protein by, for example, competitively binding to an EDIRF receptor of EDIRF-binding protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the EDIRF proteins.

In one embodiment, variants of the EDIRF protein which function as either EDIRF agonists (mimetics) or as EDIRF antagonists can be identified by screening combinatorial libraries of mutants, (e.g., truncation mutants) of the EDIRF protein for EDIRF protein agonist or antagonist activity. In one embodiment, a variegated library of EDIRF variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of EDIRF variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential EDIRF sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of EDIRF sequences therein. There are a variety of methods which can be used to produce libraries of potential

EDIRF variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential EDIRF sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of the EDIRF protein coding sequence can be used to generate a variegated population of EDIRF fragments for screening and subsequent selection of variants of an EDIRF protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an EDIRF coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the EDIRF protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of EDIRF proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recrusive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify EDIRF variants (Arkin and Yourvan (1992) *PNAS 89*:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated EDIRF library. For example, a library of expression vectors can be transfected into a cell line which ordinarily secretes EDIRF protein. Supernatants from the transfected cells are then contacted with EDIRF-responsive cells and the effect of the mutation in

EDIRF can be detected, e.g., by measuring any of a number of EDIRF-responsive cell responses. Plasmid DNA can then be recovered from the mutant EDIRF-secreting cells which score for inhibition, or alternatively, potentiation of the EDIRF-dependent response, and the individual clones further characterized.

An isolated EDIRF protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind EDIRF using standard techniques for polyclonal and monoclonal antibody preparation. The full-length EDIRF protein can be used or, alternatively, the invention provides antigenic peptide fragments of EDIRF for use as immunogens. The antigenic peptide of EDIRF comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6 and encompasses an epitope of EDIRF such that an antibody raised against the peptide forms a specific immune complex with EDIRF. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of EDIRF that are located on the surface of the protein, e.g., hydrophilic regions.

A EDIRF immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed EDIRF protein or a chemically synthesized EDIRF polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic EDIRF preparation induces a polyclonal anti-EDIRF antibody response.

Accordingly, another aspect of the invention pertains to anti-EDIRF antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as EDIRF. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind EDIRF. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of EDIRF. A monoclonal antibody

composition thus typically displays a single binding affinity for a particular EDIRF protein with which it immunoreacts.

Polyclonal anti-EDIRF antibodies can be prepared as described above by immunizing a suitable subject with an EDIRF immunogen. The anti-EDIRF antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized EDIRF. If desired, the antibody molecules directed against EDIRF can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-EDIRF antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem* .255:4980-83; Yeh et al. (1976) *PNAS* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an EDIRF immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds EDIRF.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-EDIRF monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing

hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind EDIRF, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-EDIRF antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with EDIRF to thereby isolate immunoglobulin library members that bind EDIRF. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; Dower et al. PCT International Publication No. WO 91/17271; Winter et al. PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 92/09690; Ladner et al. PCT International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J. Mol. Biol.* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *PNAS* 89:3576-3580; Garrad et al. (1991) *Bio/Technology* 9:1373-1377; Hoogenboom et al. (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas et al. (1991) *PNAS* 88:7978-7982; and McCafferty et al. *Nature* (1990) 348:552-554.

Additionally, recombinant anti-EDIRF antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in

Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *PNAS* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *PNAS* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

An anti-EDIRF antibody (e.g., monoclonal antibody) can be used to isolate EDIRF by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-EDIRF antibody can facilitate the purification of natural EDIRF from cells and of recombinantly produced EDIRF expressed in host cells. Moreover, an anti-EDIRF antibody can be used to detect EDIRF protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the EDIRF protein. Anti-EDIRF antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding EDIRF (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting

another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., EDIRF proteins, mutant forms of EDIRF, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of EDIRF in prokaryotic or eukaryotic cells. For example, EDIRF can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promotors directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in EDIRF activity assays, in EDIRF ligand binding (e.g., direct assays or competitive assays described in detail below), to generate antibodies specific for EDIRF proteins, as examples. In a preferred embodiment, an EDIRF fusion expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral

polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the EDIRF expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, EDIRF can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al.

(1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *PNAS* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to EDIRF mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, EDIRF protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or

mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding EDIRF or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) EDIRF protein. Accordingly, the invention further provides methods for producing EDIRF protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding EDIRF has been introduced) in a suitable medium such that EDIRF protein is produced. In another embodiment, the method further comprises isolating EDIRF from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which EDIRF-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous EDIRF sequences have been introduced into their genome or homologous recombinant animals in which endogenous EDIRF sequences have been altered. Such animals are useful for studying the function and/or activity of EDIRF and

for identifying and/or evaluating modulators of EDIRF activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous EDIRF gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing EDIRF-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The EDIRF cDNA sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614 can be introduced as a transgene into the genome of a non-human animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the EDIRF transgene to direct expression of EDIRF protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Patent No. 4,873,191 by Wagner et al. and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the EDIRF transgene in its genome and/or expression of EDIRF mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding EDIRF can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an EDIRF gene into which a deletion, addition or substitution has

been introduced to thereby alter, e.g., functionally disrupt, the EDIRF gene. The EDIRF gene can be a human gene (e.g., the cDNA of SEQ ID NO:1 or SEQ ID NO:5), but more preferably, is a non-human homologue of a human EDIRF gene (e.g., the cDNA of SEQ ID NO:3). For example, a mouse EDIRF gene of SEQ ID NO:3 can be used to construct a homologous recombination vector suitable for altering an endogenous EDIRF gene in the mouse genome. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous EDIRF gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous EDIRF gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous EDIRF protein). In the homologous recombination vector, the altered portion of the EDIRF gene is flanked at its 5' and 3' ends by additional nucleic acid of the EDIRF gene to allow for homologous recombination to occur between the exogenous EDIRF gene carried by the vector and an endogenous EDIRF gene in an embryonic stem cell. The additional flanking EDIRF nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced EDIRF gene has homologously recombined with the endogenous EDIRF gene are selected (see e.g., Li, E. et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec et al.; WO 91/01140 by Smithies et al.; WO 92/0968 by Zijlstra et al.; and WO 93/04169 by Berns et al.

In another embodiment, transgenic non-humans animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For

a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al. (1992) *PNAS* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. et al. (1997) *Nature* 385:810-813. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The recontructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The EDIRF nucleic acid molecules, EDIRF proteins, and anti-EDIRF antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils,

polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., an EDIRF protein or anti-EDIRF antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a

powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel™, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova

Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by

stereotactic injection (see e.g., Chen et al. (1994) *PNAS* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology), c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials); and d) methods of treatment (e.g., therapeutic and prophylactic methods as well as such methods in the context of pharmacogenetics). As described herein, an EDIRF protein of the invention has one or more of the following activities: (i) activation of an EDIRF-dependent signal transduction pathway; (ii) modulation of secretion of a non-IL-17 cytokine; (i) activation of an EDIRF-dependent signal transduction pathway; (ii) modulation of secretion of a non-IL-17 cytokine; (iii) interaction of an EDIRF protein with a non-EDIRF cell surface molecule (e.g., a proteoglycan or sulfated proteoglycan); (iv) interaction of an EDIRF protein with heparin; (v) modulation of secretion of IL-17; (vi) modulation of surface expression of a cellular adhesion molecule; (vii) modulation of a proinflammatory cytokine; (viii) modulation of a hematopoietic cytokine; (ix) modulation of the development or differentiation of an EDIRF-expressing cell; (x) modulation of the development or differentiation of a non- EDIRF-expressing cell; (xi) modulation of the homeostasis of an EDIRF-expressing cell; and (xii) modulation of the homeostasis of a non-EDIRF-expressing cell and can thus be used in, for example, (1) regulation of the cellular immune response; (2) regulation of inflammation; (3) regulation of hematopoiesis, or (4) control of differentiation or development, either *in vitro* or *in vivo*. The isolated nucleic acid molecules of the invention can be used, for example, to express EDIRF protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect EDIRF mRNA (e.g., in a biological sample) or a genetic alteration in an EDIRF gene, and to modulate EDIRF activity, as described further below. In addition, the EDIRF proteins can be used to screen drugs or compounds which modulate the EDIRF activity as well as to treat disorders characterized by insufficient or excessive production of EDIRF protein or production of

EDIRF protein forms which have decreased or aberrant activity compared to EDIRF wild type protein (e.g., inflammatory diseases such as arthritis, differentiative or developmental disorders, or immune response disorders). Moreover, soluble forms of the EDIRF protein can be used to bind other membrane-bound cytokine receptors and influence bioavailability of such a receptors cognate ligand. In addition, the anti-EDIRF antibodies of the invention can be used to detect and isolate EDIRF proteins and modulate EDIRF activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to EDIRF proteins or have a stimulatory or inhibitory effect on, for example, EDIRF expression or EDIRF activity.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of an EDIRF protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of an EDIRF receptor. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406);

(Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses an EDIRF receptor on the cell surface is contacted with a test compound and the ability of the test compound to bind to an EDIRF receptor determined. The cell, for example, can be of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to an EDIRF receptor can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the EDIRF receptor can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a test compound to interact with an EDIRF receptor without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test compound with an EDIRF receptor without the labeling of either the test compound or the receptor. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and receptor.

In a preferred embodiment, the assay comprises contacting a cell which expresses an EDIRF receptor on the cell surface with an EDIRF protein or biologically-active portion thereof, to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an EDIRF receptor, wherein determining the ability of the test compound to interact with an EDIRF receptor comprises determining the ability of the test compound to preferentially bind to the EDIRF receptor as compared to the ability of EDIRF, or a biologically active portion thereof, to bind to the receptor.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing an EDIRF target molecule with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the EDIRF target molecule. Determining the ability of the test compound to modulate the activity of EDIRF target molecule can be accomplished, for example, by determining the

ability of the EDIRF protein to bind to or interact with the EDIRF target molecule. As used herein, a "target molecule" is a molecule with which an EDIRF protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an EDIRF protein, a molecule on the surface of a second cell, or a molecule in the extracellular milieu. A EDIRF target molecule can be a non-EDIRF molecule or an EDIRF protein or polypeptide of the present invention. The target, for example, can be a membrane-bound or membrane-associated protein which facilitates signal transduction from EDIRF to an intercellular protein which has catalytic activity or to an intercellular protein which facilitates the association of downstream signaling molecules with EDIRF.

Determining the ability of the EDIRF protein to bind to or interact with an EDIRF target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the EDIRF protein to bind to or interact with an EDIRF target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (i.e. intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an EDIRF-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, an immune cell response or cell homing.

In yet another embodiment, an assay of the present invention is a cell-free assay in which an EDIRF protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the EDIRF protein or biologically active portion thereof is determined. Binding of the test compound to the EDIRF protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the EDIRF protein or biologically active portion thereof with a known compound which binds EDIRF to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an EDIRF protein, wherein determining the ability of the test compound to interact with an EDIRF protein comprises determining the ability of the test compound to preferentially bind to EDIRF or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which EDIRF protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the EDIRF protein or biologically active portion thereof is determined. Determining the ability of

the test compound to modulate the activity of EDIRF can be accomplished, for example, by determining the ability of the EDIRF protein to bind to an EDIRF target molecule by one of the methods described above for determining direct binding. Determining the ability of the EDIRF protein to bind to an EDIRF target molecule can also be accomplished using a technology such as real-time Biomolocular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore™). Changes in the optical phenomenon surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of EDIRF can be accomplished by determining the ability of the EDIRF protein to further modulate an EDIRF target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting the EDIRF protein or biologically active portion thereof with a known compound which binds EDIRF to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an EDIRF protein, wherein determining the ability of the test compound to interact with an EDIRF protein comprises determining the ability of the EDIRF protein to preferentially bind to or modulate the activity of an EDIRF target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of isolated proteins (e.g. EDIRF proteins or biologically active portions thereof or EDIRF target molecules). In the case of cell-free assays in which a membrane-bound form an isolated protein is used (e.g., an EDIRF target molecule or receptor) it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the isolated protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either EDIRF or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the

proteins, as well as to accommodate automation of the assay. Binding of a test compound to EDIRF, or interaction of EDIRF with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ EDIRF fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or EDIRF protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of EDIRF binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either EDIRF or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated EDIRF or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with EDIRF or target molecules but which do not interfere with binding of the EDIRF protein to its target molecule can be derivatized to the wells of the plate, and unbound target or EDIRF trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the EDIRF or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the EDIRF or target molecule.

In another embodiment, modulators of EDIRF expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of EDIRF mRNA or protein in the cell is determined. The level of expression of EDIRF mRNA or protein in the presence of the candidate compound is compared to the level of expression of EDIRF mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of EDIRF expression based

on this comparison. For example, when expression of EDIRF mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of EDIRF mRNA or protein expression. Alternatively, when expression of EDIRF mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of EDIRF mRNA or protein expression. The level of EDIRF mRNA or protein expression in the cells can be determined by methods described herein for detecting EDIRF mRNA or protein.

In yet another aspect of the invention, the EDIRF proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with EDIRF ("EDIRF-binding proteins" or "EDIRF-bp") and modulate EDIRF activity. Such EDIRF-binding proteins are also likely to be involved in the propagation of signals by the EDIRF proteins as, for example, downstream elements of an EDIRF-mediated signaling pathway. Alternatively, such EDIRF-binding proteins are likely to be cell-surface molecules associated with non-EDIRF expressing cells, wherein such EDIRF-binding proteins are involved in secondary cytokine production.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for EDIRF is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming an EDIRF-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with EDIRF.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an

agent identified as described herein (e.g., an EDIRF modulating agent, an antisense EDIRF nucleic acid molecule, an EDIRF-specific antibody, or an EDIRF binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the EDIRF, sequences, described herein, can be used to map the location of the EDIRF genes, respectively, on a chromosome. The mapping of the EDIRF sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, EDIRF genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the EDIRF sequences. Computer analysis of the EDIRF, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the EDIRF sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels

of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the EDIRF sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a 9o, 1p, or 1v sequence to its chromosome include *in situ* hybridization (described in Fan, Y. et al. (1990) *PNAS*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in

Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. et al. (1987) *Nature,* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the EDIRF gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The EDIRF sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the EDIRF sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The EDIRF sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that

allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:7, SEQ ID NO:8 or SEQ ID NO:9 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from EDIRF sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual: Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial EDIRF Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NOs:1, SEQ ID NO:3 or SEQ ID NO:5 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the EDIRF sequences or portions thereof, e.g., fragments derived from the noncoding

regions of SEQ ID NO:1, SEQ ID NO:3 or SEQ ID NO:5, having a length of at least 20 bases, preferably at least 30 bases.

The EDIRF sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such EDIRF probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., EDIRF primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trails are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining EDIRF protein and/or nucleic acid expression as well as EDIRF activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant EDIRF expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with EDIRF protein, nucleic acid expression or activity. For example, mutations in an EDIRF gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby phophylactically treat an individual prior to the onset of a disorder characterized by or associated with EDIRF protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of EDIRF in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of EDIRF in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting EDIRF protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes EDIRF protein such that the presence of EDIRF is detected in the biological sample. A

preferred agent for detecting EDIRF mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to EDIRF mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length EDIRF nucleic acid, such as the nucleic acid of SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to EDIRF mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting EDIRF protein is an antibody capable of binding to EDIRF protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect EDIRF mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of EDIRF mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of EDIRF protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of EDIRF genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of EDIRF protein include introducing into a subject a labeled anti-EDIRF antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a

compound or agent capable of detecting EDIRF protein, mRNA, or genomic DNA, such that the presence of EDIRF protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of EDIRF protein, mRNA or genomic DNA in the control sample with the presence of EDIRF protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of EDIRF in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting EDIRF protein or mRNA in a biological sample; means for determining the amount of EDIRF in the sample; and means for comparing the amount of EDIRF in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect EDIRF protein or nucleic acid.

2. **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant EDIRF expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with EDIRF protein, nucleic acid expression or activity such as an inflammatory or immune disorder, a differentiative or developmental disorder, or a hematopoietic disorder. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing an inflammatory or immune disease. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant EDIRF expression or activity in which a test sample is obtained from a subject and EDIRF protein or nucleic acid (e.g, mRNA, genomic DNA) is detected, wherein the presence of EDIRF protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant EDIRF expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant EDIRF expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder, such as an inflammatory or immune disorder, a differentiative or developmental disorder, or a hematopoietic disorder. Alternatively,

such methods can be used to determine whether a subject can be effectively treated with an agent for an inflammatory disease or an immune system disease. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant EDIRF expression or activity in which a test sample is obtained and EDIRF protein or nucleic acid is detected (e.g., wherein the presence of EDIRF protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant EDIRF expression or activity.)

The methods of the invention can also be used to detect genetic alterations in an EDIRF gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by aberrant development, aberrant immune responsiveness, an aberrant inflammatory response or an aberrant hematopoietic response. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding an EDIRF-protein, or the mis-expression of the EDIRF gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from an EDIRF gene; 2) an addition of one or more nucleotides to an EDIRF gene; 3) a substitution of one or more nucleotides of an EDIRF gene, 4) a chromosomal rearrangement of an EDIRF gene; 5) an alteration in the level of a messenger RNA transcript of an EDIRF gene, 6) aberrant modification of an EDIRF gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of an EDIRF gene, 8) a non-wild type level of an EDIRF-protein, 9) allelic loss of an EDIRF gene, and 10) inappropriate post-translational modification of an EDIRF-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in an EDIRF gene. A preferred biological sample is serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *PNAS* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the EDIRF-gene (see Abravaya et al. (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more

primers which specifically hybridize to an EDIRF gene under conditions such that hybridization and amplification of the EDIRF-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, 1990, Proc. Natl. Acad. Sci. USA 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et all, 1988, Bio/Technology 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an EDIRF gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in EDIRF can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in EDIRF can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential ovelapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the EDIRF gene and detect mutations by comparing the sequence of the sample EDIRF with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *PNAS* 74:560) or Sanger ((1977) *PNAS* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the EDIRF gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type EDIRF sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton et al (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in EDIRF cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on an EDIRF sequence, e.g., a wild-type EDIRF sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme,

and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in EDIRF genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control EDIRF nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential

hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an EDIRF gene.

Furthermore, any cell type or tissue in which EDIRF is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of EDIRF (e.g., activation of an EDIRF-dependent signal transduction pathway; modulation of secretion of a non-IL-17 cytokine; modulation of surface expression of a cellular adhesion molecule; modulation of a proinflammatory cytokine; or modulation of a hematopoietic cytokine) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase EDIRF gene expression, protein levels, or upregulate EDIRF activity, can be monitored in clinical trails of subjects exhibiting decreased EDIRF gene expression, protein levels, or downregulated EDIRF activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease EDIRF gene expression, protein levels, or downregulate EDIRF activity, can be monitored in clinical trails of subjects exhibiting increased EDIRF gene expression, protein levels, or upregulated EDIRF activity. In such clinical trials, the expression or activity of EDIRF and, preferably, other genes that have been implicated in, for example, an immune response disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including EDIRF, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule)

which modulates EDIRF activity (e.g., identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on immune response disorders, developmental disorder, or hematopoietic disorder, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of EDIRF and other genes implicated in the immune response disorders, developmental disorder, or hematopoietic disorder, respectively. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of EDIRF or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of an EDIRF protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the EDIRF protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the EDIRF protein, mRNA, or genomic DNA in the pre-administration sample with the EDIRF protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of EDIRF to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of EDIRF to lower levels than detected, i.e. to decrease the effectiveness of the agent. According to such an embodiment, EDIRF expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

C. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant EDIRF expression or activity. With regards to both prophylactic and

therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the EDIRF molecules of the present invention or EDIRF modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. **Prophylactic Methods**

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant EDIRF expression or activity, by administering to the subject an agent which modulates EDIRF expression or at least one EDIRF activity. Subjects at risk for a disease which is caused or contributed to by aberrant EDIRF expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the EDIRF aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of EDIRF aberrancy, for example, an EDIRF agonist or EDIRF antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the present invention are further discussed in the following subsections.

2. **Therapeutic Methods**

Another aspect of the invention pertains to methods of modulating EDIRF expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of EDIRF protein activity associated with the cell. An agent that modulates EDIRF protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of an EDIRF protein, a peptide, an EDIRF peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more EDIRF protein activity. Examples of such stimulatory agents include active

EDIRF protein and a nucleic acid molecule encoding EDIRF that has been introduced into the cell. In another embodiment, the agent inhibits one or more EDIRF protein activity. Examples of such inhibitory agents include antisense EDIRF nucleic acid molecules and anti-EDIRF antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an EDIRF protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) EDIRF expression or activity. In another embodiment, the method involves administering an EDIRF protein or nucleic acid molecule as therapy to compensate for reduced or aberrant EDIRF expression or activity.

Stimulation of EDIRF activity is desirable in situations in which EDIRF is abnormally downregulated and/or in which increased EDIRF activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant immune responsiveness. Another example of such a situation is where the subject has a inflammatory disease (e.g., arthritis or psoriasis). Yet another example of such a situation is where a subject has a disorder characterized by an aberrant hematopoietic response. Yet another example of such a situation is where a subject has a disorder characterized by aberrant differentiation or development.

3. Pharmacogenomics

The EDIRF molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on EDIRF activity (e.g., EDIRF gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g, immune response disorders or developmental disorders) associated with aberrant EDIRF activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer an EDIRF molecule or EDIRF modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with an EDIRF molecule or EDIRF modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, M., *Clin Exp Pharmacol Physiol*, 1996, 23(10-11) :983-985 and Linder, M.W., *Clin Chem*, 1997, 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drugs target is known (e.g., an EDIRF protein or EDIRF receptor of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., an EDIRF molecule or EDIRF modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an EDIRF molecule or EDIRF modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are hereby incorporated by reference.

EXAMPLES

Example 1: Isolation And Characterization of Human EDIRF cDNAs

In this example, the isolation and characterization of the gene encoding human EDIRF (also referred to as "TANGO 89") is described.

Isolation of a first human EDIRF cDNA (human EDIRF I), clone jthua012a05

A human EDIRF cDNA was isolated from a human fetal lung cDNA library. To construct the library, three micrograms of poly A^+ RNA were isolated from human fetal lung tissue, reverse transcribed, and used to synthesize a cDNA library using the Superscript cDNA Synthesis kit™ (Gibco BRL; Gaithersburg, MD). Complementary DNA was directionally cloned into the expression plasmid pMET7 using the SalI and NotI sites in the polylinker to construct a plasmid library. Transformants were picked and amplified for single-pass sequencing. Additionally, human fetal lung cDNA was ligated into the SalI/NotI sites of the ZIPLOX™ vector (Gibco BRL) for construction of a lambda phage cDNA library.

Clone jthua012a05 was identified from the above-described cDNA library using the following method. First, each sequence was checked to determine if it was a bacterial, ribosomal, or mitochondrial contaminant. Such sequences were excluded from the subsequent analysis. Second, sequence artifacts, such as vector and repetitive elements, were masked and/or removed from each sequence. Third, the remaining sequences were searched against a copy of the GenBank nucleotide database using the BLASTN™ program (BLASTN 1.3MP: Altschul et al., *J. Mol. Bio.* 215:403, 1990). Fourth, the sequences were analyzed against a non-redundant protein database with the BLASTX™ program, which translates a nucleic acid sequence in all six frames and compares it against available protein databases (BLASTX 1.3MP:Altschul et al., *supra*). This protein database is a combination of the Swiss-Prot, PIR, and NCBI GenPept protein databases.

The original first pass sequence of clone jthua012a05 showed homology to interleukin-17 using the BLASTX™ program. The nucleotide sequence and predicted amino acid sequence are shown in Figure 1A (corresponding to SEQ ID NO:1 and SEQ ID NO:2, respectively.) The COOH-terminus of this human EDIRF protein (corresponding to amino acids 95-180 of the predicted amino acid sequence) shows 41% identity with IL-17 (see Figure 3). This human EDIRF protein contains a signal sequence from amino acids 1-22.

A BLASTN™ search of the EST database revealed the following ESTs having significant homology to clone jthua012a05:

EST Database hits	Species	Base Pairs Covered	% Identity	Coding?
Accession # AA033733	human	60-511	99	yes
Accession # AA443286	human	718-321	99	yes
Accession # AA680405	human	706-310	99	yes
Accession # W74664	human	713-259	91	yes
Accession # AA044549	mouse	20-517	87	yes
Accession # W74558	human	39-411	99	yes
Accession # W87101	mouse	60-489	78	yes
Accession # W89980	mouse	90-504	76	yes
Accession # W83241	mouse	12-497	78	yes

Isolation of the murine EDIRF cDNA (murine EDIRF I)

The gene encoding human IL-17 was initially used to conduct a database search of published expressed sequence tag (EST) sequences using the BLAST program TBLASTN (Washington University version 2.0). All matched sequences with a score of 70 or better (BLOSUM62 substitution matrix, default gap penalties of -10 for insertion and -10 for extension, no pre-filtering of the query sequence) were selected for further analysis. Matched sequences were utilized to search a database of known proteins using BLASTX (parameters were set the same as for the TBLASTN search described above). Sequences corresponding to known proteins (those having from 90-100% amino acid identity to the query matched sequence) were discarded from the analysis. Five mouse EST sequences were identified in the EST database (Accession Nos.: AA015563, W83241, W87101, AA044549, and W89980). Each sequence had been deposited by Washington University as part of the WashU-HHMI Mouse EST project and the corresponding clones were available through the Research Genetics IMAGE Consortium (IMAGE clone #468019, #405946, #407352, #475876, and #407493, respectively). Clone #407352 was subsequently amplified and fully sequenced. The full-length nucleotide sequence of murine EDIRF is presented in Figure 2 in addition to the predicted amino acid sequence. The nucleotide and amino acid sequences of Figure 2 correspond to those of SEQ ID NO:3 and SEQ ID NO:4, respectively.

The COOH-terminus of murine EDIRF (corresponding to amino acids 95-180 of the predicted amino acid sequence) shows 38% identity with IL-17 (see Figure 3). This murine EDIRF protein contains a signal sequence from amino acids 1-22.

Isolation of a second human EDIRF cDNA (human EDIRF II)

In addition to the five murine EST sequences described above, two additional human nucleotide sequences were published in the EST database (Accession Nos.: W74664 and W74558) which had a significant degree of homology to human IL-17. The sequences had been deposited by Washington University as part of the WashU-Merck Mouse EST project and the corresponding clone was available through the Research Genetics IMAGE Consortium (IMAGE clone #344649). When this clone was isolated and fully sequenced, however, it showed no homology to either the database entries or to the known sequence for IL-17. Four oligonucleotides were therefore designed from the published human EST sequence (oDH53, oDH54, oDH55 and oDH56). These oligos were used as PCR primers in an attempt to amplify corresponding cDNA sequences from several libraries selected a potential sources of human EDIRF sequences. Two oligos were used as forward primers: oDH53 and oDH54, and two as reverse primers: oDH55 and oDH56.

oDH53	5' GCC CAG CTG AGG AAC AGT 3' (SEQ ID NO:17)
oDH54	5' GAG CTG GCC CAG AGA AAG 3' (SEQ ID NO:18)
oDH55	5' ACA GGA ACC TGG CTG AAC 3' (SEQ ID NO:19)
oDH56	5' AGC CCA CAG CGA TGG TCT 3' (SEQ ID NO:20)

Two microliters of each of the following cDNA libraries were used in the PCR amplification reactions:

1. Clontech™ lung (lambda gt10)
2. ZIPLOX™ fetal lung*
3. ZIPLOX bronchial epithelial cell*
4. Clontech™ T Cell (lambda gt11)
5. Clontech™ Jurkat (lambda gt11)
6. Clontech™ lymphocyte (lambda gt11)
7. Stratagene™ lung (lambda ZAP™)
8. ZIPLOX™ HL60 stimulated cell line*
9. ZIPLOX™ HUVEC*
10. ZIPLOX™ HMVEC*
11. monocyte (lambda ZAP™)*

*ZIPLOX™ cDNA libraries were constructed utilizing mRNA isolated from the respective cell or tissue sources using the ZIPLOX™ cDNA library construction kit (Gibco BRL)

Each possible primer combination was used with the following PCR profile:

94 C° - 4' followed by

94° C - 30"
55° C - 30"
72° C - 30"
for 30 cycles.

Expected size of amplification products:

oDH53 & oDH55	240 bp
oDH53 & oDH56	320 bp
oDH54 & oDH55	220 bp
oDH54 & oDH56	300 bp

Various nucleic acid fragments were identified by ethidium bromide staining of PCR products electrophoresed through agarose gels, however, none were or the predicted size. Using the primer combination oDH53 and oDH56, there were distinctly-sized PCR products from three of the cDNA libraries: ZIPLOX™ fetal lung, ZIPLOX™ bronchial epithelial cell, and Stratagene™ lung. The PCR products from each library were identical and were 220 bp, 400 bp, and 600 bp in size.

These reaction products were each purified using standard conditions detailed in the Qiagen QIAquick PCR Purification kit™. Two microliters of the purified reaction products were used in a second PCR reaction with the primers oDH54 & oDH55 (both interior to oDH53 & oDH56). The secondary reaction performed utilizing the primary PCR product from the ZIPLOX™ fetal lung library was of the predicted size (220 bp).

This 220 bp fragment was excised from the gel and the PCR product sequenced. In addition it was subcloned using Invitrogen's TA Cloning kit™. The nucleotide sequence of the PCR product had homology to the published human EST sequences. It was therefore utilized to isolate additional clones in an attempt to obtain the full-length sequence of this second human EDIRF cDNA.

The 220 bp fragment was labeled using Stratagene's Prime-It RT kit™ and hybridized with approximately 10^6 clones of the ZIPLOX™ fetal lung library, plated on 20 NZY plates. Eight primary positive clones were identified in the first round of screening. These were eluted and replated on fresh plates, and five of the eight showed strong positive hybridization signals in the secondary screening. Isolated plaques from the positive phage were then eluted (four of the five positives required tertiary purification), an aliquot of each (25 microliters) was combined with 100 microliters DH10B ZIP™ cells to excise the pZL1 plasmid. One of the resulting bacterial colonies was amplified and plasmid DNA was purified by Qiagen™ mini-prep kit and digested with SalI and NotI to determine the insert size. The 1.5kB insert was ultimately

sequenced and the nucleotide sequence is shown in Figure 1B (the nucleotide sequence corresponds to SEQ ID NO:5). This nucleotide sequence corresponds to human EDIRF II which has significant homology to human EDIRF I and includes additional 5' nucleotide sequence not present in human EDIRF I. The predicted amino acid sequence of human EDIRF II is also shown in Figure 1B (corresponding to SEQ ID NO:6). The nucleotide sequence contains an open reading frame from nucleotide 309-911, encoding a predicted protein of 261 amino acids. The COOH-terminus of this human EDIRF protein (corresponding to amino acids 176-261 of the predicted amino acid sequence) shows 41% identity with IL-17 (see Figure 3).

Example 2: Distribution of EDIRF mRNA In Human and Murine Tissues

Northern Blot Analysis

The expression of EDIRF was analyzed using Northern blot hybridization. For analysis of human EDIRF, two primers pDH140RP01 [5' TTA CCA TTT CCA TCT TCC TGG 3'] (SEQ ID NO:10) and pDH140FP01 [5' TGC AGG TGC AGC CCA CAG 3'] (SEQ ID NO:11) were used to amplify a 504 base pair (bp) fragment for use as a probe (corresponding to nucleotides 100-603 of SEQ ID NO:1 and to nucleotides 403-906 of SEQ ID NO:5). For murine EDIRF, a vector primer adjacent to the insert was used in conjunction with primer pDH93FP01 [5' GCA TGG TGA AGG GAT TCA CGC 3'] (SEQ ID NO:12) to amplify a 491 bp fragment covering 470 nucleotides of the 5' end of the EDIRF gene (corresponding to nucleotides 448 through 756 of SEQ ID NO:3). The probe DNAs were radioactively labeled with ^{32}P-dCTP using the Prime-It kit™ (Stratagene, La Jolla, CA) according to the instructions of the supplier. Filters containing mRNA (human MTNI and MTNII and murine embryo MTN from Clontech, Palo Alto, CA;) were probed in ExpressHyb™ hybridization solution (Clontech) and washed at high stringency according to manufacturer's recommendations.

Low level expression of an approximately 1.0 kb transcript was observed in murine adult lung. In addition, a high molecular weight signal (hybridization between 4.4 and 10kb) was observed in adult murine brain and, to a lower level, in skeletal muscle which is believed to be unrelated to the 1.0kb EDIRF signal. On a Northern blot of fetal mRNA, a developmentally regulated transcript of ~1.0 kb was observed. The signal peaked in mRNA isolated from day 15 embryos, was significantly lower at both day 11 and day 17, and was absent at day 7. No other hybridization signal was observed (see below for in situ hybridization results.)

The mRNA expression of human EDIRF differed from that observed for murine EDIRF. Varying levels of two bands at ~5.0kb and <1kb were observed in all tissues

(spleen, thymus, prostate, testis, ovary, small intestine, colon, peripheral blood leukocytes, heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas.) In addition, a faint transcript at ~1.0 kb was seen in testis, heart, and liver. Furthermore, the Northern blots of adult mouse and human mRNAs took significantly longer to generate an exposure for, suggesting that the highest levels of this transcript are embryonic.

***In situ* Hybridization Analysis**

For *in situ* hybridization analysis of murine EDIRF, 10μm sagittal sections of fresh frozen embryonic day 13.5 and postnatal day 1.5 B6 mice, as well as 8μm cross sections of adult B6 mouse lung and heart tissues were used for hybridization. Sections were postfixed with 4% formaldehyde in DEPC-treated 1X phosphate- buffered saline at room temperature for 10 minutes before being rinsed twice in DEPC-treated 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections were rinsed in DEPC-treated 2X SSC (1X SSC is 0.15M NaCl plus 0.015M sodium citrate). Tissue was dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

The hybridization was performed using a ^{35}S-radiolabeled cRNA probe from the following DNA sequence (corresponding to nucleotides 246-510 of SEQ ID NO:3):

GAATGGAAGAGTATGAGCGGAACCTTGGGGAGATGGTGGCCCAGCTGAGGAACAGCTC
CGAGCCAGCCAAGAAGAAATGTGAAGTCAATCTACAGCTGTGGTTGTCCAACAAGAGG
AGCCTGTCCCCATGGGGCTACAGCATCAACCACGACCCCAGCCGCATCCCTGCGGACT
TGCCCGAGGCGCGGTGCCTATGTTTGGGTTGCGTGAATCCCTTCACCATGCAGGAGGA
CCGTAGCATGGTGAGCGTGCCAGTGTTCAGCCA (SEQ ID NO:13)

Tissues were incubated with probe (approximately 5 X 10^7 cpm/ml) in the presence of a solution containing 600 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1 X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 h at 55°C. After hybridization, slides were washed with 2 X SSC. Sections were then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10ug of RNase A per ml for 30

minutes, and finally in TNE for 10 minutes. Slides were then rinsed with 2 X SSC at room temp, washed with 2 X SSC at 50°C for 1 hour, washed with 0.2 X SSC at 55°C for 1 hour, and 0.2 X SSC at 60°C for 1 hour. Sections were then dehydrated rapidly through serial ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR™ scientific imaging film for 6 days at room temperature.

Following a 6 day film exposure signal could be seen only in the embryonic day 13.5 sections. Signal was apparent in the diaphragm and skeletal/cartilage structure or muscle structure throughout the embryo, and was readily observed in the developing spinal column and sternum. Signal was also observed at similar intensity in the tongue, developing jaw area, tissue surrounding the eye, tissue along the back of the neck, a thin layer just below the skin throughout most of the embryo, and in the developing appendages, all in a pattern similar to that of muscle. Signal was not observed in the developing cranium or large bone of the legs. This absence of signal suggests that the EDIRF signal most likely originated from developing muscle rather that bone or cartilage.

No signal was seen in the postnatal day 1.5 or adult lung and heart tissue sections. The postnatal day 1.5 tissue sections contain similar structures as the embryonic day 13.5 sections (diaphragm, sternum, spinal column, etc.) leading to the conclusion that EDIRF expression drops below detectable limits between embryonic day 13.5 and postnatal day 1.5, consistent with the results obtained from the analysis of embryonic expression using the Northern blots.

Example 3: Characterization of Murine EDIRF I Epitope-Tagged Protein

Expression of Murine EDIRF I:flag

A murine EDIRF I flag epitope-tagged protein (mEDIRF:flag) vector was constructed by PCR followed by ligation into an expression vector, pMET stop. The full-length open-reading frame was PCR amplified using a 5' primer incorporating a Kozak sequence upstream of and including the initiator methionine and a 3' primer including the nucleotide sequence encoding the DYKDDDDK flag epitope (SEQ ID NO:21) followed by a termination codon. The primer sequences are shown below:

5' primer (muT89F): 5' AAA AAA GAA TTC GCC GCC ATG GAC TGG CCG CAC 3' (SEQ ID NO:22)

3' primer (muT89RF): 5' TCC TCT GTC GAC TCA CTT GTC GTC GTC GTC CTT GTA GTC GAA GAT GCA GGT GCA 3' (SEQ ID NO:23)

The sequenced DNA construct was transiently transfected into HEK 293T cells in 150mM plates using Lipofectamine (GIBCO/BRL, Gaithersburg MD) according to the manufacturer's protocol. Seventy-two hours post-transfection, the serum-free conditioned medium (OptiMEM, GIBCO /BRL) was harvested and spun.

N-terminal sequencing of mEDIRF:flag

The 293 conditioned medium was electroblotted onto a PVDF membrane (Novex) after separation by SDS-PAGE on 4-20% gradient gels and probed with M2 anti-flag polyclonal antibody diluted 1:500 (Sigma Chemical Co., St. Louis MO) followed by horseradish peroxidase conjugated sheep anti-mouse antibody diluted 1:5000 (Amersham Pharmacia Biotech, Inc., Piscataway NJ) developed with chemiluminescent reagents (Renaissance, DuPont-NEN Research, Boston MA) and exposed to autoradiography film (BioMax MR2 film, Eastman Kodak Co., Rochester, NY). Flag immunoreactivity appeared as two pairs of bands that migrated on SDS-PAGE between the 17 and 30 kDa markers (Multimark, Novex).

Samples of each of the four mEDIRF:flag bands that were bound to PVDF membrane after SDS-PAGE and electroblotting were subjected to N-terminal amino acid sequence analysis using Edman-based chemistry protein sequencing on a PE Applied Biosystems Model 494 sequencer (Applied Biosystems, Inc., Foster City CA). The amino acid residues were analyzed by HPLC using a Spherogel micro PTH 3-micron column and identified by separation and peak height as compared to standards. As shown below in Table 1, one band (band C) gave no interpretable signal and the remaining three have the same N-terminal sequence. As all these bands share the same N-terminus and the same C-terminus (flag), the primary structure of the different molecular weight bands is likely to be the same. The different forms may arise through non-proteolytic post-translational modification, for example, by differences in N-glycosylation at a predicted N-glycosylation site (Asn75 of SEQ ID NO:3). The sequence obtained confirms the deduced amino acid sequence encompassing amino acids 23 to 33 shown in Figure 2.

Determination of mEDIRF:flag protein molecular weight

Nine hundred fifty mL of the 293 conditioned medium was passed over a sulphopropyl cation exchange resin (Poros HS, PE Biosystems Division of The Perkin-Elmer Corp., Norwalk CT) equilibrated in 20mM MES pH 6.0 and eluted with a gradient of 0 - 1.0 M NaCl in the same buffer. Samples of unbound material and fractions collected during the elution were analyzed by silver stained SDS-PAGE and SDS-PAGE/flag western blot as above. No mEDIRF:flag immunoreactive material was detected in the unbound sample, whereas intense immunoreactivity was observed in fractions eluted over the gradient corresponding to between 0.55M and 0.8M NaCl. The molecular weight of each of the four bands on silver-stained SDS-PAGE were measured relative to protein molecular weight standards. The calculated molecular weights in kDa are summarized below in Table 1.

Table 1. N-terminal sequence and MW of mEDIRF:flag

	N-terminal sequence	Molecular weight
band A	RNTKGKRKGQG	25
band B	RNTKGKRKGQG	23
band C	Not determined	20
band D	RNTKGKRKGQG	19

Because the purity of cation exchange purified mEDIRF:flag was less than 50%, a second purification step was employed to eliminate non-immunoreactive contaminants. mEDIRF:flag fractions obtained by cation exchange were pooled, then fractionated over a quaternized polyethyleneimine anion exchange resin (Poros HQ, PE Biosystems) equilibrated in 20mM Tris pH 8.0 and eluted over a gradient of 0 - 1.0 M NaCl in the same buffer. Samples of unbound material and of fractions collected during the elution were analyzed by silver stained SDS-PAGE and SDS-PAGE/flag western blot as above. No mEDIRF:flag immunoreactive material was detected in the unbound sample, whereas intense immunoreactivity was observed in fractions eluted over the gradient corresponding to between 0.3M and 0.4M NaCl. The purity of mEDIRF:flag

after sequential cation and anion exchange chromatography was more than 90% as judged by silver stained SDS-PAGE.

Example 4: Localization of Murine EDIRF:flag

To analyze cell-associated mEDIRF:flag, HEK 293T cells were transfected as above in multiple wells of 6-well tissue culture plates. After 72 hours, cells were washed with PBS and then incubated with 0.05% trypsin in Hank's medium or in a PBS control. After 0, 3 and 20 minutes, wells were scraped and cells pelleted by centrifugation. Cell pellets were boiled in a reducing SDS-PAGE sample buffer and analyzed by SDS-PAGE/flag western blot. The PBS control cell pellets contained significant mEDIRF:flag immunoreactivity. Cells treated with trypsin for 3 or 20 minutes exhibited more than 3-fold reduction in immunoreactivity, suggesting that most of cell-associated murine EDIRF is exposed to trypsin and thus on the cell surface rather than within the plasma membrane.

To analyze the release of mEDIRF:flag from the cell surface, HEK 293 cells were transfected as above in multiple wells of 6-well tissue culture plates. After 72 hours, cells were washed with PBS and then incubated with heparin (1mg/mL or 0.01mg/mL) or PBS as control. After 60 minutes at 4°C, supernatants and cells were harvested by centrifugation and boiled in SDS-PAGE sample buffer. SDS-PAGE/flag western blot analysis showed that addition of soluble heparin results in an increase in mEDIRF:flag immunoreactivity in the supernatant and a corresponding decrease in the pellet relative to controls. Both concentrations of heparin were equally effective. These data suggest that mEDIRF:flag is associated with the surface of the cell through association with sulfated proteoglycan, probably heparin. Thus, EDIRF is predicted to be cell-associated, suggesting that EDIRF may function primarily as a localized modulator.

Example 5: Binding of EDIRF I to Mouse Tissues

Preparation of AP:EDIRF I

A mouse EDIRF I alkaline phosphatase N-terminal fusion protein (AP:mEDIRF) vector was constructed by PCR followed by ligation into the AP-Tag3 vector (Tartaglia, L., et al., 1995. Cell 83:1263-1271). The full-length open-reading frame of murine EDIRF I was PCR amplified using a 5' primer incorporating a BglII restriction site prior to the nucleotides encoding the first amino acids of the mature murine EDIRF I protein (RNTK) and a 3' primer including a XhoI restriction site

immediately following the termination codon of murine EDIRF I. Thus the open reading frame of the complete construct includes the complete sequence of human placental alkaline phosphatase, including the signal peptide, followed by the mature murine EDIRF I DNA sequence. The primer sequences are shown below:

5' primer: 5' TTT TTA GAT CTC GGA ACA CCA AAG GCA AAA G 3' (SEQ ID NO:24)

3' primer: 5' TTT TGT CGA CTC AGA AGA TGC AGG TGC AAC 3' (SEQ ID NO:25)

The sequenced DNA constructs were transiently transfected into HEK 293T cells in 150mM plates using Lipofectamine (GIBCO/BRL) according to the manufacturer's protocol. Seventy-two hours post-transfection, the serum-free conditioned media (OptiMEM, GIBCO/BRL) were harvested, spun and filtered. Alkaline phosphatase activity in conditioned media was quantitated using an enzymatic assay kit (Phospha-Light, Tropix Inc.) according to the manufacturer's instructions.

When alkaline phosphatase fusion protein concentrations below 2 nM were observed, conditioned medium was concentrated by centrifugation using a 30 kDa cut-off membrane. Conditioned medium samples before and after concentration were analyzed by SDS-PAGE followed by Western blot using anti-human alkaline phosphatase antibodies diluted 1:250 (Genzyme Corp., Cambridge MA) and detected by chemiluminescence. A band at 90 kDa was observed in concentrated supernatants of AP:mEDIRF transfections. Conditioned medium samples were adjusted to 10 % fetal calf serum and stored at 4°C.

Screening of tissue sections for AP:mEDIRF binding

Supernatant containing AP:mEDIRF was used to screen tissue sections essentially as described by Cheng and Flanagan (1994) *Cell* 79:157-168. Briefly, fresh frozen sections (8 μm) were prepared and rinsed in HBHA (Hank's balanced salt solution supplemented with 20 mM Hepes, pH 7.0, 0.05% BSA, 0.1% sodium azide). Tissue sections were subsequently incubated with supernatant containing AP:mEDIRF or alkaline phosphatase at a concentration of 5nM for one hour at room temperature. After the incubation, tissue sections were washed six times in HBHA, fixed (60% acetone, 3% formaldehyde, 20 mM Hepes, pH 7.5), washed three times in HBS (20 mM

Hepes, pH 7.5, 150 mM NaCl), then heated for 30 min at 65°C to inactivate endogenous alkaline phosphatase activity. Bound AP:mEDIRF was detected by developing sections in BCIP/NBT substrate solution (100 mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM MgCl, 0.17 mg/ml BCIP and 0.33 mg/ml NBT).

Using this procedure, tissue sections of mouse embryos (day 14.5 of prenatal development) and newborn mice (postnatal day 1.5) were screened for EDIRF I binding. AP:mEDIRF but not alkaline phosphatase alone bound strongly to all tissue structures. In the presence of increasing concentrations of heparin (30U - 1000 U/ml), the overall binding of AP:mEDIRF generally decreased, although binding of AP:mEDIRF to certain structures such as fetal liver, skeletal muscle and cartilage of the embryo sections was retained. This suggests that EDIRF I may be binding to heparin or to heparin-like molecules in the majority of fetal and newborn mouse tissues. The binding characteristics of a human EDIRF I alkaline phosphatase N-terminal fusion protein were similar to AP:mEDIRF.

Example 6: Expression of Human EDIRF I

Construction of a Human EDIRF I expression vector

Human EDIRF I was PCR amplified from jthua012a05 using the following 5' and 3' PCR primers:

5' (B91): GATCGATCGATCCCATGGACTGGCCTCACAA (SEQ ID NO:26)

3' (B92): TGACTGACTGACTGTCGACTCACTACTTGTCATCGTCATCCTTAT AGTCGAAGATGCAGGTGCAG (SEQ ID NO:27)

The resulting PCR product encoded an NCO I site in frame with the initiating MET of human EDIRF at the 5' end and a fused flag tag (GACTATAAGGATGACGATGACAAG) (SEQ ID NO:28) followed by two stop codons (TAGTGA) and a HindIII site at the 3' end. The resulting product is as follows:

```
      +1                              M  D   W  P  H  N   L  L  F
                                    NcoI
                                    ~~~~~~
      1  GATCGATCGA TCTCTAGAGC CACCATGGAC TGGCCTCACA ACCTGCTGTT
         CTAGCTAGCT AGAGATCTCG GTGGTACCTG ACCGGAGTGT TGGACGACAA

      +1   L  L  T   I  S  I  F   L  G  L   G  Q  P   R  S  P
     51  TCTTCTTACC ATTTCCATCT TCCTGGGGCT GGGCCAGCCC AGGAGCCCCA
         AGAAGAATGG TAAAGGTAGA AGGACCCCGA CCCGGTCGGG TCCTCGGGGT
```

```
  +1 K  S  K  R  K  G  Q  G  R  P  G  P  L  A  P  G  P
101  AAAGCAAGAG GAAGGGGCAA GGGCGGCCTG GGCCCCTGGC CCCTGGCCCT
     TTTCGTTCTC CTTCCCCGTT CCCGCCGGAC CCGGGGACCG GGGACCGGGA

  +1 H  Q  V  P  L  D  L  V  S  R  M  K  P  Y  A  R  M
151  CACCAGGTGC CACTGGACCT GGTGTCACGG ATGAAACCGT ATGCCCGCAT
     GTGGTCCACG GTGACCTGGA CCACAGTGCC TACTTTGGCA TACGGGCGTA

  +1  E  E  Y  E  R  N  I  E  E  M  V  A  Q  L  R  N
201  GGAGGAGTAT GAGAGGAACA TCGAGGAGAT GGTGGCCCAG CTGAGGAACA
     CCTCCTCATA CTCTCCTTGT AGCTCCTCTA CCACCGGGTC GACTCCTTGT

  +1 S  S  E  L  A  Q  R  K  C  E  V  N  L  Q  L  W  M
251  GCTCAGAGCT GGCCCAGAGA AAGTGTGAGG TCAACTTGCA GCTGTGGATG
     CGAGTCTCGA CCGGGTCTCT TTCACACTCC AGTTGAACGT CGACACCTAC

  +1 S  N  K  R  S  L  S  P  W  G  Y  S  I  N  H  D  P
301  TCCAACAAGA GGAGCCTGTC TCCCTGGGGC TACAGCATCA ACCACGACCC
     AGGTTGTTCT CCTCGGACAG AGGGACCCCG ATGTCGTAGT TGGTGCTGGG

  +1  S  R  I  P  V  D  L  P  E  A  R  C  L  C  L  G
351  CAGCCGTATC CCCGTGGACC TGCCGGAGGC ACGGTGCCTG TGTCTGGGCT
     GTCGGCATAG GGGCACCTGG ACGGCCTCCG TGCCACGGAC ACAGACCCGA

  +1 C  V  N  P  F  T  M  Q  E  D  R  S  M  V  S  V  P
401  GTGTGAACCC CTTCACCATG CAGGAGGACC GCAGCATGGT GAGCGTGCCG
     CACACTTGGG GAAGTGGTAC GTCCTCCTGG CGTCGTACCA CTCGCACGGC

  +1 V  F  S  Q  V  P  V  R  R  R  L  C  P  P  P  P  R
451  GTGTTCAGCC AGGTTCCTGT GCGCCGCCGC CTCTGCCCGC CACCGCCCCG
     CACAAGTCGG TCCAAGGACA CGCGGCGGCG GAGACGGGCG GTGGCGGGGC

  +1  T  G  P  C  R  Q  R  A  V  M  E  T  I  A  V  G
501  CACAGGGCCT TGCCGCCAGC GCGCAGTCAT GGAGACCATC GCTGTGGGCT
     GTGTCCCGGA ACGGCGGTCG CGCGTCAGTA CCTCTGGTAG CGACACCCGA

  +1 C  T  C  I  F  D  Y  K  D  D  D  D  K  *  *
                                                 HindIII
                                                 ~~~~~~
551  GCACCTGCAT CTTCGACTAT AAGGATGACG ATGACAAGTA GTGAAAGCTT
     CGTGGACGTA GAAGCTGATA TTCCTACTGC TACTGTTCAT CACTTTCGAA

601  AGCTAGCTA
     TCGATCGAT
```

This PCR product was cut with NCO I and HindIII and cloned into pEF/myc/cyto (a pShooter™ vector from Invitrogen Corp.) cut with the same restriction enzymes. The resulting plasmid, pEF_IL17h, was grown in ElectroMAX DH10B™ competent cells (GIBCO/BRL), and purified using the QIAFilter system (QIAGEN, Inc.)

Expression of Human EDIRF I

293T cells were grown to 80% confluency, then split into T75 flasks at 6 x 10^6 cells per flask in DMEM/F12, 10%FBS culture media and incubated overnight at 37°C. Next day, the cells were washed twice with 1x PBS and DMEM/F12 supplemented with 10mM Hepes, Insulin-Transferrin-Selenium-X supplement (GIBCO), and 2mM glutamine. After one hour incubation at 37°C in the serum-free medium, the cultures were transfected with pEF_IL17h in FuGENE™ 6 Transfection Reagent (Boehringer Mannheim Corp., Indianapolis IN). The FuGENE/DNA suspension, per T75 flask, was prepared as follows: 45 µl FuGENE was added to 750 µl DMEM/F12 supplemented media, then vortexed. Eleven µg of pEF_IL17h (at a concentration of 0.5 to 1µg/µl) was then added and incubated for 30 minutes before adding to the cells. The supernatant was harvested after 72 hours incubation.

Example 7: Purification of Human EDIRF I

Human EDIRF I was purified from approximately 125 ml of media pooled from three 72-hour supernatants obtained as described above.

The media was pumped over a freshly prepared fast flow S column (30 ml bed volume; Pharmacia LKB) at a flow rate of 5 ml/min. After all of the media had been loaded, the column was washed with 50 mM Tris pH 7.5 until the absorbance at 280 nm returned to near baseline. A linear gradient from 0-1M NaCl (in 50 mM Tris pH 7.5) was started and developed over 60 min (at a flow rate of 5 ml/min. Absorbance was monitored at 280 nm and 5 ml fractions were collected. The fractions were analyzed by gel electrophoresis (4-20 % SDS-PAGE) and fractions containing protein at approximately the expected size were electrophoresed (4-20 % SDS-PAGE), transferred to a PVDF membrane and stained with Coomassie Blue.

Example 8: N-terminal sequencing of Human EDIRF I

SDS-PAGE protein bands obtained as above were excised and sequenced on an ABI 494 sequencer (Applied Biosystems, Inc.), essentially following the procedure as described above in Example 3 for the N-terminal sequencing of murine EDIRF I from which the following sequence was determined:

RSPKSKRKGQ (SEQ ID NO:29)

The sequence obtained confirms the deduced amino acid sequence encompassing amino acids 23 to 32 shown in Figure 1A.

Example 9: Effect of EDIRF on Early Development

Preparation of murine EDIRF I RNA

The template for the murine EDIRF *in vitro* transcription reaction was prepared from the DNA construct for the murine EDIRF flag epitope-tagged protein as described above in Example 3 inserted in a pCS2 vector, which was then linearized using AscI. Capped RNA was synthesized using SP6 RNA polymerase from the linearized plasmid using mMESSAGE mMACHINE kit (Ambion, Austin, TX) according to the manufacturer's instructions. Phenol:chloroform (1:1)-extracted RNA was precipitated with isopropanol. Prior to injection into embryos, the RNA was tested for translatability *in vitro* in the presence of ^{35}S-methionine using a rabbit reticulocyte lysate system (Promega, Madison WI).

Murine EDIRF I RNA injection into Xenopus embryos

Xenopus embryos were obtained by in vitro fertilization, dejellied in 2% cysteine HCl (pH 7.6), washed thoroughly in Modified Ringers solution, and incubated at 15-25° C. Embryos were transferred to injection solution (Modified Ringers solution containing 5% Ficoll) prior to injections. 100-500 pg of murine EDIRF I RNA was injected into each blastomere at the 2-cell stage. Embryos were transferred to water from the injection solution after approximately 6 hours and grown until the appropriate stage.

Examination two days later of embryos injected with Tango 89 RNA showed an overexpression phenotype. These results suggest that EDIRF I has a dramatic effect on early tissue differentiation.

Analyses of embryos for EDIRF I or EDIRF II expression

Embryos for Western analysis are frozen and then lyzed in buffer containing 50mM TrisCl pH7.5, 0.5M urea, 2% NP40, 1mMPMSF and 5% 2-ME. Proteins are separated by gel electrophoresis and Western blotted according to standard methods. Immunodetection of protein is carried out using a BM chemiluminescence Western blotting kit according to the manufacterer's instructions (Boehringer Mannheim). Diluted anti-EDIRF I or anti-EDIRF II rabbit antiserum will be used as primary antibody.

Embryos for histological examination are fixed in 10% Zn formalin overnight, embedded in paraffin and stained with H&E by standard procedures.

Embryos for immunohistochemistry are fixed in 100% methanol at –20°C, embedded in paraffin and sectioned at 8um as above and processed by standard procedures using various primary antibodies and horseradish peroxidase-conjugated secondary antibodies for detection.

Xenopus cap assay

100-500 pg of EDIRF I or EDIRF II RNA are injected into the animal pole of each of the 2 blastomeres at the 2-cell stage. Animal caps from uninjected or injected embryos will be explanted at stage 9 and cultured in 0.1x Modified Ringers containing 0.1% BSA and 50ug/ml gentamycin. Alternatively, uninjected animal caps will be incubated in 10-100ng/ml of purified EDIRF I or EDIRF II protein. Animal caps are cultured until control embryos have reached stage 19-20. Poly A+ RNA is extracted from lysed animal caps and whole embryos using mAP paper (Amersham, IL) as described by Amaravadi and King (1994) *Oncogene* 9:971-974. RT-PCR is performed on these samples using gene-specific primers and appropriate annealing temperatures and the products analyzed by gel electrophoresis.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% homologous to a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 500 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, wherein the fragment comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614;

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, wherein the nucleic acid molecule

hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof under stringent conditions; and

f) a nucleic acid molecule which hybridizes under stringent conditions to a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof.

2. The isolated nucleic acid molecule of claim 1 which is selected from the group consisting of:

a) a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, or SEQ ID NO:5, or a complement thereof, under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 65% homologous to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614;

d) a polypeptide which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to a nucleotide sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the DNA insert of the plasmid deposited with ATCC as Accession Number 98614; and

e) a polypeptide comprising an amino acid sequence which is at least 60% homologous to the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the polypeptide encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the polypeptide encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614;

b) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98613, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number 98614, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 8 in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule in claim 1 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of claim 1 in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for EDIRF activity.

21. A method of modulating the activity of a polypeptide of claim 8 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8 comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

Figure 1A

```
                                                                  M  D          2
GTCGACCCACGCGTCCGCTGGGGTTCCAGGCGGGCAGCAGCTGCAGGCTGACCTTGCAGCTTGGCGGA ATG GAC      74

 W  P  H  N  L  L  F  L  L  T  I  S  I  F  L  G  L  G  Q  P    22
TGG CCT CAC AAC CTG CTG TTT CTT CTT ACC ATT TCC ATC TTC CTG GGG CTG GGC CAG CCC   134

 R  S  P  K  S  K  R  K  G  Q  G  R  P  G  P  L  A  P  G  P    42
AGG AGC CCC AAA AGC AAG AGG AAG GGG CAA GGG CGG CCT GGG CCC CTG GCC CCT GGC CCT   194

 H  Q  V  P  L  D  L  V  S  R  M  K  P  Y  A  R  M  E  E  Y    62
CAC CAG GTG CCA CTG GAC CTG GTG TCA CGG ATG AAA CCG TAT GCC CGC ATG GAG GAG TAT   254

 E  R  N  I  E  E  M  V  A  Q  L  R  N  S  S  E  L  A  Q  R    82
GAG AGG AAC ATC GAG GAG ATG GTG GCC CAG CTG AGG AAC AGC TCA GAG CTG GCC CAG AGA   314

 K  C  E  V  N  L  Q  L  W  M  S  N  K  R  S  L  S  P  W  G   102
AAG TGT GAG GTC AAC TTG CAG CTG TGG ATG TCC AAC AAG AGG AGC CTG TCT CCC TGG GGC   374

 Y  S  I  N  H  D  P  S  R  I  P  V  D  L  P  E  A  R  C  L   122
TAC AGC ATC AAC CAC GAC CCC AGC CGT ATC CCC GTG GAC CTG CCG GAG GCA CGG TGC CTG   434

 C  L  G  C  V  N  P  F  T  M  Q  E  D  R  S  M  V  S  V  P   142
TGT CTG GGC TGT GTG AAC CCC TTC ACC ATG CAG GAG GAC CGC AGC ATG GTG AGC GTG CCG   494

 V  F  S  Q  V  P  V  R  R  R  L  C  P  P  P  P  R  T  G  P   162
GTG TTC AGC CAG GTT CCT GTG CGC CGC CGC CTC TGC CCG CCA CCG CCC CGC ACA GGG CCT   554

 C  R  Q  R  A  V  M  E  T  I  A  V  G  C  T  C  I  F  *      181
TGC CGC CAG CGC GCA GTC ATG GAG ACC ATC GCT GTG GGC TGC ACC TGC ATC TTC TGA       611

ATCACCTGGCCCAGAAGCCAGGCCAGCAGCCCGAGACCATCCTCCTTGCACCTTTGTGCCAAGAAAGGCCTATGAAAAG  690

TAAACACTGACTTTTGAAAGCAAAAAAAAAAAAAAAAAAAAAGGGCGGCC                              738
```

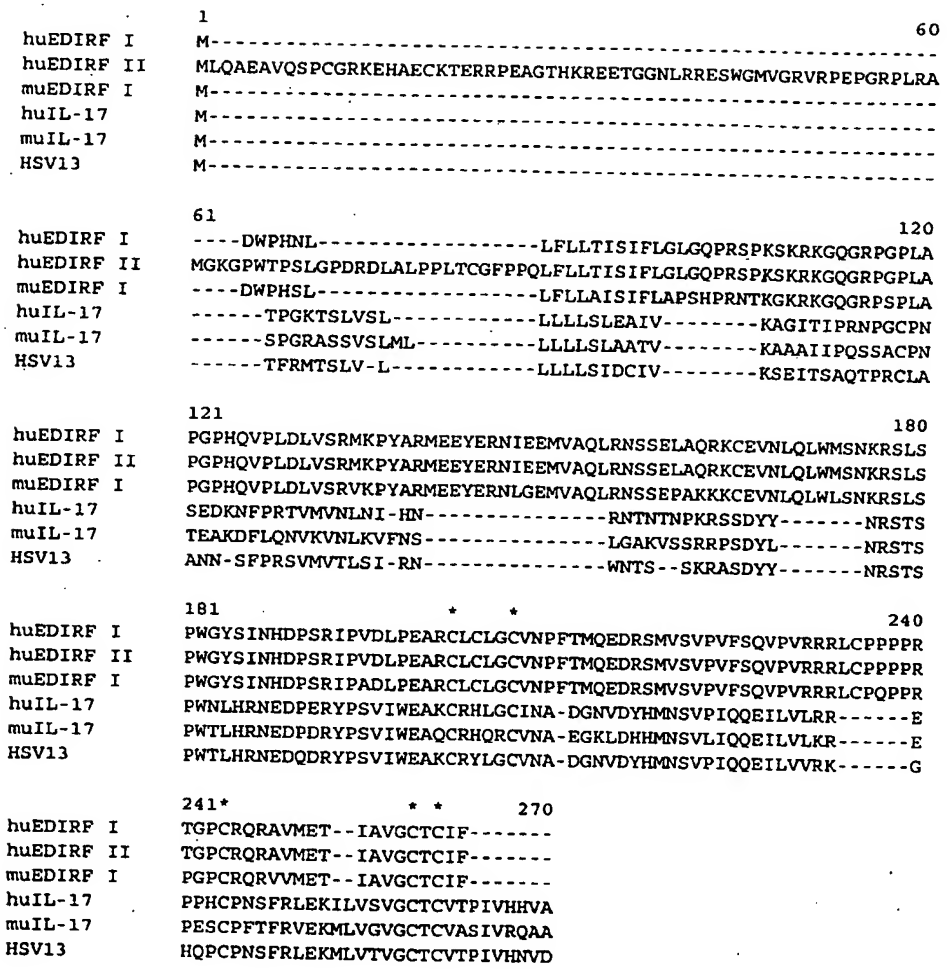
Figure 1B

```
CCACGCGTCCGGTCATGCAAGAGGTGTGTGGGTTATGAGAGCCCAAGAAAGTCCAGAGGCATCCCTAAGAAAGCGATGC   79

                                                    M  L  Q  A  E  A  V      7
TAAAGCTGACACGAGGGGTGGGAAGGTGTCCAGGAGGTGAAGAGGCAGA ATG CTG CAG GCA GAG GCT GTG   149

 Q  S  P  C  G  R  K  E  H  A  E  C  K  T  E  R  R  P  E  A   27
CAA AGT CCC TGT GGC AGG AAG GAG CAT GCT GAG TGT AAG ACT GAG AGA AGA CCA GAG GCT  209

 G  T  H  K  R  E  E  T  G  G  N  L  R  R  E  S  W  G  M  V   47
GGA ACA CAC AAG AGA GAG GAA ACA GGG GGT AAT CTG AGG AGG GAG AGC TGG GGG ATG GTA  269

 G  R  V  R  P  E  P  G  R  P  L  R  A  M  G  K  G  P  W  T   67
GGC AGA GTC AGA CCT GAG CCT GGG AGA CCC CTA AGA GCA ATG GGA AAG GGA CCA TGG ACC  329

 P  S  L  G  P  D  R  D  L  A  L  P  P  L  T  C  G  F  P  P   87
CCA AGT CTT GGC CCA GAT AGA GAC TTG GCT CTC CCA CCT TTG ACC TGT GGC TTT CCT CCC  389

 Q  L  F  L  L  T  I  S  I  F  L  G  L  G  Q  P  R  S  P  K  107
CAG CTG TTT CTT CTT ACC ATT TCC ATC TTC CTG GGG CTG GGC CAG CCC AGG AGC CCC AAA  449

 S  K  R  K  G  Q  G  R  P  G  P  L  A  P  G  P  H  Q  V  P  127
AGC AAG AGG AAG GGG CAA GGG CGG CCT GGG CCC CTG GCC CCT GGC CCT CAC CAG GTG CCA  509

 L  D  L  V  S  R  M  K  P  Y  A  R  M  E  E  Y  E  R  N  I  147
CTG GAC CTG GTG TCA CGG ATG AAA CCG TAT GCC CGC ATG GAG GAG TAT GAG AGG AAC ATC  569

 E  E  M  V  A  Q  L  R  N  S  S  E  L  A  Q  R  K  C  E  V  167
GAG GAG ATG GTG GCC CAG CTG AGG AAC AGC TCA GAG CTG GCC CAG AGA AAG TGT GAG GTC  629

 N  L  Q  L  W  M  S  N  K  R  S  L  S  P  W  G  Y  S  I  N  187
AAC TTG CAG CTG TGG ATG TCC AAC AAG AGG AGC CTG TCT CCC TGG GGC TAC AGC ATC AAC  689

 H  D  P  S  R  I  P  V  D  L  P  E  A  R  C  L  C  L  G  C  207
CAC GAC CCC AGC CGT ATC CCC GTG GAC CTG CCG GAG GCA CGG TGC CTG TGT CTG GGC TGT  749

 V  N  P  F  T  M  Q  E  D  R  S  M  V  S  V  P  V  F  S  Q  227
GTG AAC CCC TTC ACC ATG CAG GAG GAC CGC AGC ATG GTG AGC GTG CCG GTG TTC AGC CAG  809

 V  P  V  R  R  R  L  C  P  P  P  P  R  T  G  P  C  R  Q  R  247
GTT CCT GTG CGC CGC CGC CTC TGC CCG CCA CCG CCC CGC ACA GGG CCT TGC CGC CAG CGC  869

 A  V  M  E  T  I  A  V  G  C  T  C  I  F  *                    262
GCA GTC ATG GAG ACC ATC GCT GTG GGC TGC ACC TGC ATC TTC TGA                    914

ATCACCTGGCCCAGAAGCCAGGCCAGCAGCCCGAGACCATCCTCCTTGCACCTTTGTGCCAAGAAAGGCCTATGAAAAG  993

TAAACACTGACTTTTGAAAGCAAGACTTTGGGTTTGCTTTTTGTGGCTTCGAAGTCAAAACCAGGGTCAGGGTAGGAGG  1072

TTCCTGGGAGGTAGATTTCAGATTAATATAAAGGAGGATTTTTCATCACTCTGAGTTCTCCAATACTGAGGTGGCCTAT  1151

TAGAAAATAATGTGTTCACCTCTACCACCTAGGGCCATCATGTGTTTCCTTTATTTGTGACAGAGTGAGACTGTCAAAA  1230

AAAAAAAAAAAAAAAAAAAAAAAAAAAAA                                                  1259
```

Figure 2

```
                                                                           M      1
GAATTCGGCACGATGACGGCATCTGGGGTTCCTGGCGGGTGGCAGCTGCGGGCCTGCCGCCTGACTTGGTGGG ATG     76

 D   W   P   H   S   L   L   F   L   L   A   I   S   I   F   L   A   P   S   H    21
GAC TGG CCG CAC AGC CTG CTC TTC CTC CTG GCC ATC TCC ATC TTC CTG GCG CCA AGC CAC  136

 P   R   N   T   K   G   K   R   K   G   Q   G   R   P   S   P   L   A   P   G    41
CCC CGG AAC ACC AAA GGC AAA AGA AAA GGG CAA GGG AGG CCC AGT CCC TTG GCC CCT GGG  196

 P   H   Q   V   P   L   D   L   V   S   R   V   K   P   Y   A   R   M   E   E    61
CCT CAT CAG GTG CCG CTG GAC CTG GTG TCT CGA GTA AAG CCC TAC GCT CGA ATG GAA GAG  256

 Y   E   R   N   L   G   E   M   V   A   Q   L   R   N   S   S   E   P   A   K    81
TAT GAG CGG AAC CTT GGG GAG ATG GTG GCC CAG CTG AGG AAC AGC TCC GAG CCA GCC AAG  316

 K   K   C   E   V   N   L   Q   L   W   L   S   N   K   R   S   L   S   P   W   101
AAG AAA TGT GAA GTC AAT CTA CAG CTG TGG TTG TCC AAC AAG AGG AGC CTG TCC CCA TGG  376

 G   Y   S   I   N   H   D   P   S   R   I   P   A   D   L   P   E   A   R   C   121
GGC TAC AGC ATC AAC CAC GAC CCC AGC CGC ATC CCT GCG GAC TTG CCC GAG GCG CGG TGC  436

 L   C   L   G   C   V   N   P   F   T   M   Q   E   D   R   S   M   V   S   V   141
CTA TGT TTG GGT TGC GTG AAT CCC TTC ACC ATG CAG GAG GAC CGT AGC ATG GTG AGC GTG  496

 P   V   F   S   Q   V   P   V   R   R   R   L   C   P   Q   P   P   R   P   G   161
CCA GTG TTC AGC CAG GTG CCG GTG CGC CGC CGC CTC TGT CCT CAA CCT CCT CGC CCT GGG  556

 P   C   R   Q   R   V   V   M   E   T   I   A   V   G   C   T   C   I   F   *   181
CCC TGC CGC CAG CGT GTC GTC ATG GAG ACC ATC GCT GTG GGT TGC ACC TGC ATC TTC TGA  616

GCCAACCACCAACCCGGTGGCCTCTGCAACAACCCTCCCTCCCTGCACCCACTGTGACCCTCAAGGCTGATAAACAGTA  695

AACGCTGTTCTTTGTAAAGGAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTCCGCGGCCGC                     756
```

Figure 3


1 60
huEDIRF I M
huEDIRF II MLQAEAVQSPCGRKEHAECKTERRPEAGTHKREETGGNLRRESWGMVGRVRPEPGRPLRA
muEDIRF I M
huIL-17 M
muIL-17 M
HSV13 M
61 120
huEDIRF I ----DWPHNL----------LFLLTISIFLGLGQPRSPKSKRKGQGRPGPLA
huEDIRF II MGKGPWTPSLGPDRDLALPPLTCGFPPQLFLLTISIFLGLGQPRSPKSKRKGQGRPGPLA
muEDIRF I ----DWPHSL----------LFLLAISIFLAPSHPRNTKGKRKGQGRPSPLA
huIL-17 ------TPGKTSLVSL----------LLLLSLEAIV--------KAGITIPRNPGCPN
muIL-17 ------SPGRASSVSLML----------LLLLSLAATV--------KAAAIIPQSSACPN
HSV13 ------TFRMTSLV-L----------LLLLSIDCIV--------KSEITSAQTPRCLA
121 180
huEDIRF I PGPHQVPLDLVSRMKPYARMEEYERNIEEMVAQLRNSSELAQRKCEVNLQLWMSNKRSLS
huEDIRF II PGPHQVPLDLVSRMKPYARMEEYERNIEEMVAQLRNSSELAQRKCEVNLQLWMSNKRSLS
muEDIRF I PGPHQVPLDLVSRVKPYARMEEYERNLGEMVAQLRNSSEPAKKKCEVNLQLWLSNKRSLS
huIL-17 SEDKNFPRTVMVNLNI-HN----------RNTNTNPKRSSDYY-------NRSTS
muIL-17 TEAKDFLQNVKVNLKVFNS----------LGAKVSSRRPSDYL-------NRSTS
HSV13 ANN-SFPRSVMVTLSI-RN----------WNTS--SKRASDYY-------NRSTS
181 * * 240
huEDIRF I PWGYSINHDPSRIPVDLPEARCLCLGCVNPFTMQEDRSMVSVPVFSQVPVRRRLCPPPPR
huEDIRF II PWGYSINHDPSRIPVDLPEARCLCLGCVNPFTMQEDRSMVSVPVFSQVPVRRRLCPPPPR
muEDIRF I PWGYSINHDPSRIPADLPEARCLCLGCVNPFTMQEDRSMVSVPVFSQVPVRRRLCPPPPR
huIL-17 PWNLHRNEDPERYPSVIWEAKCRHLGCINA-DGNVDYHMNSVPIQQEILVLRR------E
muIL-17 PWTLHRNEDPDRYPSVIWEAQCRHQRCVNA-EGKLDHHMNSVLIQQEILVLKR------E
HSV13 PWTLHRNEDQDRYPSVIWEAKCRYLGCVNA-DGNVDYHMNSVPIQQEILVVRK------G
241* * * 270
huEDIRF I TGPCRQRAVMET--IAVGCTCIF-------
huEDIRF II TGPCRQRAVMET--IAVGCTCIF-------
muEDIRF I PGPCRQRVVMET--IAVGCTCIF-------
huIL-17 PPHCPNSFRLEKILVSVGCTCVTPIVHHVA
muIL-17 PESCPFTFRVEKMLVGVGCTCVASIVRQAA
HSV13 HQPCPNSFRLEKMLVTVGCTCVTPIVHNVD

SEQUENCE LISTING

<110> Millennum Biotherapeutics Inc.

<120> NOVEL EMBRYO-DERIVED INTERLEUKIN RELATED FACTOR MOLECULES AND USES THEREFOR

<130> MEI-007CPPC

<140>
<141>

<150> US 08/944,890
<151> 1997-12-19

<160> 29

<170> PatentIn Ver. 2.0

<210> 1
<211> 738
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (69)..(608)

<400> 1
gtcgacccac gcgtccgctg gggttccagg cgggcagcag ctgcaggctg accttgcagc 60

ttggcgga atg gac tgg cct cac aac ctg ctg ttt ctt ctt acc att tcc 110
Met Asp Trp Pro His Asn Leu Leu Phe Leu Leu Thr Ile Ser
1 5 10

atc ttc ctg ggg ctg ggc cag ccc agg agc ccc aaa agc aag agg aag 158
Ile Phe Leu Gly Leu Gly Gln Pro Arg Ser Pro Lys Ser Lys Arg Lys
15 20 25 30

ggg caa ggg cgg cct ggg ccc ctg gcc cct ggc cct cac cag gtg cca 206
Gly Gln Gly Arg Pro Gly Pro Leu Ala Pro Gly Pro His Gln Val Pro
35 40 45

ctg gac ctg gtg tca cgg atg aaa ccg tat gcc cgc atg gag gag tat 254
Leu Asp Leu Val Ser Arg Met Lys Pro Tyr Ala Arg Met Glu Glu Tyr
50 55 60

gag agg aac atc gag gag atg gtg gcc cag ctg agg aac agc tca gag 302
Glu Arg Asn Ile Glu Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu
65 70 75

ctg gcc cag aga aag tgt gag gtc aac ttg cag ctg tgg atg tcc aac 350
Leu Ala Gln Arg Lys Cys Glu Val Asn Leu Gln Leu Trp Met Ser Asn
80 85 90

```
aag agg agc ctg tct ccc tgg ggc tac agc atc aac cac gac ccc agc   398
Lys Arg Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser
 95                 100                 105                 110

cgt atc ccc gtg gac ctg ccg gag gca cgg tgc ctg tgt ctg ggc tgt   446
Arg Ile Pro Val Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys
                115                 120                 125

gtg aac ccc ttc acc atg cag gag gac cgc agc atg gtg agc gtg ccg   494
Val Asn Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro
            130                 135                 140

gtg ttc agc cag gtt cct gtg cgc cgc cgc ctc tgc ccg cca ccg ccc   542
Val Phe Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Pro Pro Pro
        145                 150                 155

cgc aca ggg cct tgc cgc cag cgc gca gtc atg gag acc atc gct gtg   590
Arg Thr Gly Pro Cys Arg Gln Arg Ala Val Met Glu Thr Ile Ala Val
    160                 165                 170

ggc tgc acc tgc atc ttc tgaatcacct ggcccagaag ccaggccagc         638
Gly Cys Thr Cys Ile Phe
175                 180

agcccgagac catcctcctt gcacctttgt gccaagaaag gcctatgaaa agtaaacact 698

gacttttgaa agcaaaaaaa aaaaaaaaaa aagggcggcc                       738


<210> 2
<211> 180
<212> PRT
<213> Homo sapiens

<400> 2
Met Asp Trp Pro His Asn Leu Leu Phe Leu Leu Thr Ile Ser Ile Phe
  1               5                  10                  15

Leu Gly Leu Gly Gln Pro Arg Ser Pro Lys Ser Lys Arg Lys Gly Gln
             20                  25                  30

Gly Arg Pro Gly Pro Leu Ala Pro Gly Pro His Gln Val Pro Leu Asp
         35                  40                  45

Leu Val Ser Arg Met Lys Pro Tyr Ala Arg Met Glu Glu Tyr Glu Arg
     50                  55                  60

Asn Ile Glu Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu Leu Ala
 65                  70                  75                  80

Gln Arg Lys Cys Glu Val Asn Leu Gln Leu Trp Met Ser Asn Lys Arg
                 85                  90                  95
```

```
Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser Arg Ile
            100                 105                 110

Pro Val Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys Val Asn
        115                 120                 125

Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro Val Phe
    130                 135                 140

Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Pro Pro Pro Arg Thr
145                 150                 155                 160

Gly Pro Cys Arg Gln Arg Ala Val Met Glu Thr Ile Ala Val Gly Cys
                165                 170                 175

Thr Cys Ile Phe
            180


<210> 3
<211> 756
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (74)..(613)

<400> 3
gaattcggca cgatgacggc atctggggtt cctggcgggt ggcagctgcg ggcctgccgc 60

ctgacttggt ggg atg gac tgg ccg cac agc ctg ctc ttc ctc ctg gcc    109
               Met Asp Trp Pro His Ser Leu Leu Phe Leu Leu Ala
                 1               5                   10

atc tcc atc ttc ctg gcg cca agc cac ccc cgg aac acc aaa ggc aaa   157
Ile Ser Ile Phe Leu Ala Pro Ser His Pro Arg Asn Thr Lys Gly Lys
         15                  20                  25

aga aaa ggg caa ggg agg ccc agt ccc ttg gcc cct ggg cct cat cag   205
Arg Lys Gly Gln Gly Arg Pro Ser Pro Leu Ala Pro Gly Pro His Gln
     30                  35                  40

gtg ccg ctg gac ctg gtg tct cga gta aag ccc tac gct cga atg gaa   253
Val Pro Leu Asp Leu Val Ser Arg Val Lys Pro Tyr Ala Arg Met Glu
 45                  50                  55                  60

gag tat gag cgg aac ctt ggg gag atg gtg gcc cag ctg agg aac agc   301
Glu Tyr Glu Arg Asn Leu Gly Glu Met Val Ala Gln Leu Arg Asn Ser
                 65                  70                  75

tcc gag cca gcc aag aag aaa tgt gaa gtc aat cta cag ctg tgg ttg   349
Ser Glu Pro Ala Lys Lys Lys Cys Glu Val Asn Leu Gln Leu Trp Leu
             80                  85                  90
```

```
tcc aac aag agg agc ctg tcc cca tgg ggc tac agc atc aac cac gac   397
Ser Asn Lys Arg Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp
         95                  100                 105

ccc agc cgc atc cct gcg gac ttg ccc gag gcg cgg tgc cta tgt ttg   445
Pro Ser Arg Ile Pro Ala Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu
    110                 115                 120

ggt tgc gtg aat ccc ttc acc atg cag gag gac cgt agc atg gtg agc   493
Gly Cys Val Asn Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser
125                 130                 135                 140

gtg cca gtg ttc agc cag gtg ccg gtg cgc cgc cgc ctc tgt cct caa   541
Val Pro Val Phe Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Gln
                145                 150                 155

cct cct cgc cct ggg ccc tgc cgc cag cgt gtc gtc atg gag acc atc   589
Pro Pro Arg Pro Gly Pro Cys Arg Gln Arg Val Val Met Glu Thr Ile
            160                 165                 170

gct gtg ggt tgc acc tgc atc ttc tgagccaacc accaacccgg tggcctctgc  643
Ala Val Gly Cys Thr Cys Ile Phe
        175                 180

aacaaccctc cctccctgca cccactgtga ccctcaaggc tgataaacag taaacgctgt 703

tctttgtaaa ggaaaaaaaa aaaaaaaaaa aaaaaaaaaa tttccgcggc cgc        756
```

<210> 4
<211> 180
<212> PRT
<213> Mus musculus

```
<400> 4
Met Asp Trp Pro His Ser Leu Leu Phe Leu Leu Ala Ile Ser Ile Phe
  1               5                  10                  15

Leu Ala Pro Ser His Pro Arg Asn Thr Lys Gly Lys Arg Lys Gly Gln
             20                  25                  30

Gly Arg Pro Ser Pro Leu Ala Pro Gly Pro His Gln Val Pro Leu Asp
         35                  40                  45

Leu Val Ser Arg Val Lys Pro Tyr Ala Arg Met Glu Glu Tyr Glu Arg
     50                  55                  60

Asn Leu Gly Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu Pro Ala
 65                  70                  75                  80

Lys Lys Lys Cys Glu Val Asn Leu Gln Leu Trp Leu Ser Asn Lys Arg
                 85                  90                  95
```

```
Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser Arg Ile
            100                 105                 110

Pro Ala Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys Val Asn
        115                 120                 125

Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro Val Phe
    130                 135                 140

Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Gln Pro Pro Arg Pro
145                 150                 155                 160

Gly Pro Cys Arg Gln Arg Val Val Met Glu Thr Ile Ala Val Gly Cys
                165                 170                 175

Thr Cys Ile Phe
            180


<210> 5
<211> 1259
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (129)..(911)

<400> 5
ccacgcgtcc ggtcatgcaa gaggtgtgtg ggttatgaga gcccaagaaa gtccagaggc 60

atccctaaga aagcgatgct aaagctgaca cgaggggtgg gaaggtgtcc aggaggtgaa 120

gaggcaga atg ctg cag gca gag gct gtg caa agt ccc tgt ggc agg aag  170
         Met Leu Gln Ala Glu Ala Val Gln Ser Pro Cys Gly Arg Lys
           1               5                  10

gag cat gct gag tgt aag act gag aga aga cca gag gct gga aca cac   218
Glu His Ala Glu Cys Lys Thr Glu Arg Arg Pro Glu Ala Gly Thr His
 15                  20                  25                  30

aag aga gag gaa aca ggg ggt aat ctg agg agg gag agc tgg ggg atg   266
Lys Arg Glu Glu Thr Gly Gly Asn Leu Arg Arg Glu Ser Trp Gly Met
                35                  40                  45

gta ggc aga gtc aga cct gag cct ggg aga ccc cta aga gca atg gga   314
Val Gly Arg Val Arg Pro Glu Pro Gly Arg Pro Leu Arg Ala Met Gly
            50                  55                  60

aag gga cca tgg acc cca agt ctt ggc cca gat aga gac ttg gct ctc   362
Lys Gly Pro Trp Thr Pro Ser Leu Gly Pro Asp Arg Asp Leu Ala Leu
        65                  70                  75
```

```
cca cct ttg acc tgt ggc ttt cct ccc cag ctg ttt ctt ctt acc att   410
Pro Pro Leu Thr Cys Gly Phe Pro Pro Gln Leu Phe Leu Leu Thr Ile
     80                  85                  90

tcc atc ttc ctg ggg ctg ggc cag ccc agg agc ccc aaa agc aag agg   458
Ser Ile Phe Leu Gly Leu Gly Gln Pro Arg Ser Pro Lys Ser Lys Arg
 95                 100                 105                 110

aag ggg caa ggg cgg cct ggg ccc ctg gcc cct ggc cct cac cag gtg   506
Lys Gly Gln Gly Arg Pro Gly Pro Leu Ala Pro Gly Pro His Gln Val
                115                 120                 125

cca ctg gac ctg gtg tca cgg atg aaa ccg tat gcc cgc atg gag gag   554
Pro Leu Asp Leu Val Ser Arg Met Lys Pro Tyr Ala Arg Met Glu Glu
            130                 135                 140

tat gag agg aac atc gag gag atg gtg gcc cag ctg agg aac agc tca   602
Tyr Glu Arg Asn Ile Glu Glu Met Val Ala Gln Leu Arg Asn Ser Ser
        145                 150                 155

gag ctg gcc cag aga aag tgt gag gtc aac ttg cag ctg tgg atg tcc   650
Glu Leu Ala Gln Arg Lys Cys Glu Val Asn Leu Gln Leu Trp Met Ser
    160                 165                 170

aac aag agg agc ctg tct ccc tgg ggc tac agc atc aac cac gac ccc   698
Asn Lys Arg Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro
175                 180                 185                 190

agc cgt atc ccc gtg gac ctg ccg gag gca cgg tgc ctg tgt ctg ggc   746
Ser Arg Ile Pro Val Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly
                195                 200                 205

tgt gtg aac ccc ttc acc atg cag gag gac cgc agc atg gtg agc gtg   794
Cys Val Asn Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val
            210                 215                 220

ccg gtg ttc agc cag gtt cct gtg cgc cgc cgc ctc tgc ccg cca ccg   842
Pro Val Phe Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Pro Pro
        225                 230                 235

ccc cgc aca ggg cct tgc cgc cag cgc gca gtc atg gag acc atc gct   890
Pro Arg Thr Gly Pro Cys Arg Gln Arg Ala Val Met Glu Thr Ile Ala
    240                 245                 250

gtg ggc tgc acc tgc atc ttc tgaatcacct ggcccagaag ccaggccagc      941
Val Gly Cys Thr Cys Ile Phe
255                 260

agcccgagac catcctcctt gcacctttgt gccaagaaag gcctatgaaa agtaaacact 1001

gacttttgaa agcaagactt tgggtttgct ttttgtggct tcgaagtcaa aaccagggtc 1061

agggtaggag gttcctggga ggtagatttc agattaatat aaaggaggat ttttcatcac 1121
```

tctgagttct ccaatactga ggtggcctat tagaaaataa tgtgttcacc tctaccacct 1181

agggccatca tgtgtttcct ttatttgtga cagagtgaga ctgtcaaaaa aaaaaaaaaa 1241

aaaaaaaaaa aaaaaaaa 1259

<210> 6
<211> 261
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Leu Gln Ala Glu Ala Val Gln Ser Pro Cys Gly Arg Lys Glu His
  1               5                  10                  15

Ala Glu Cys Lys Thr Glu Arg Arg Pro Glu Ala Gly Thr His Lys Arg
             20                  25                  30

Glu Glu Thr Gly Gly Asn Leu Arg Arg Glu Ser Trp Gly Met Val Gly
         35                  40                  45

Arg Val Arg Pro Glu Pro Gly Arg Pro Leu Arg Ala Met Gly Lys Gly
     50                  55                  60

Pro Trp Thr Pro Ser Leu Gly Pro Asp Arg Asp Leu Ala Leu Pro Pro
 65                  70                  75                  80

Leu Thr Cys Gly Phe Pro Pro Gln Leu Phe Leu Leu Thr Ile Ser Ile
                 85                  90                  95

Phe Leu Gly Leu Gly Gln Pro Arg Ser Pro Lys Ser Lys Arg Lys Gly
            100                 105                 110

Gln Gly Arg Pro Gly Pro Leu Ala Pro Gly Pro His Gln Val Pro Leu
        115                 120                 125

Asp Leu Val Ser Arg Met Lys Pro Tyr Ala Arg Met Glu Glu Tyr Glu
    130                 135                 140

Arg Asn Ile Glu Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu Leu
145                 150                 155                 160

Ala Gln Arg Lys Cys Glu Val Asn Leu Gln Leu Trp Met Ser Asn Lys
                165                 170                 175

Arg Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser Arg
            180                 185                 190

Ile Pro Val Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys Val
        195                 200                 205

Asn Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro Val
    210                 215                 220
```

```
Phe Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Pro Pro Pro Arg
225                 230                 235                 240

Thr Gly Pro Cys Arg Gln Arg Ala Val Met Glu Thr Ile Ala Val Gly
                245                 250                 255

Cys Thr Cys Ile Phe
            260
```

<210> 7
<211> 540
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(540)

<400> 7

```
atg gac tgg cct cac aac ctg ctg ttt ctt ctt acc att tcc atc ttc   48
Met Asp Trp Pro His Asn Leu Leu Phe Leu Leu Thr Ile Ser Ile Phe
  1               5                  10                  15

ctg ggg ctg ggc cag ccc agg agc ccc aaa agc aag agg aag ggg caa   96
Leu Gly Leu Gly Gln Pro Arg Ser Pro Lys Ser Lys Arg Lys Gly Gln
             20                  25                  30

ggg cgg cct ggg ccc ctg gcc cct ggc cct cac cag gtg cca ctg gac   144
Gly Arg Pro Gly Pro Leu Ala Pro Gly Pro His Gln Val Pro Leu Asp
         35                  40                  45

ctg gtg tca cgg atg aaa ccg tat gcc cgc atg gag gag tat gag agg   192
Leu Val Ser Arg Met Lys Pro Tyr Ala Arg Met Glu Glu Tyr Glu Arg
     50                  55                  60

aac atc gag gag atg gtg gcc cag ctg agg aac agc tca gag ctg gcc   240
Asn Ile Glu Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu Leu Ala
 65                  70                  75                  80

cag aga aag tgt gag gtc aac ttg cag ctg tgg atg tcc aac aag agg   288
Gln Arg Lys Cys Glu Val Asn Leu Gln Leu Trp Met Ser Asn Lys Arg
                85                  90                  95

agc ctg tct ccc tgg ggc tac agc atc aac cac gac ccc agc cgt atc   336
Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser Arg Ile
            100                 105                 110

ccc gtg gac ctg ccg gag gca cgg tgc ctg tgt ctg ggc tgt gtg aac   384
Pro Val Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys Val Asn
        115                 120                 125
```

```
ccc ttc acc atg cag gag gac cgc agc atg gtg agc gtg ccg gtg ttc   432
Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro Val Phe
    130                 135                 140

agc cag gtt cct gtg cgc cgc cgc ctc tgc ccg cca ccg ccc cgc aca   480
Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Pro Pro Pro Arg Thr
145                 150                 155                 160

ggg cct tgc cgc cag cgc gca gtc atg gag acc atc gct gtg ggc tgc   528
Gly Pro Cys Arg Gln Arg Ala Val Met Glu Thr Ile Ala Val Gly Cys
                165                 170                 175

acc tgc atc ttc                                                  540
Thr Cys Ile Phe
            180
```

<210> 8
<211> 540
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (1)..(540)

<400> 8

```
atg gac tgg ccg cac agc ctg ctc ttc ctc ctg gcc atc tcc atc ttc   48
Met Asp Trp Pro His Ser Leu Leu Phe Leu Leu Ala Ile Ser Ile Phe
  1               5                  10                  15

ctg gcg cca agc cac ccc cgg aac acc aaa ggc aaa aga aaa ggg caa   96
Leu Ala Pro Ser His Pro Arg Asn Thr Lys Gly Lys Arg Lys Gly Gln
             20                  25                  30

ggg agg ccc agt ccc ttg gcc cct ggg cct cat cag gtg ccg ctg gac   144
Gly Arg Pro Ser Pro Leu Ala Pro Gly Pro His Gln Val Pro Leu Asp
         35                  40                  45

ctg gtg tct cga gta aag ccc tac gct cga atg gaa gag tat gag cgg   192
Leu Val Ser Arg Val Lys Pro Tyr Ala Arg Met Glu Glu Tyr Glu Arg
     50                  55                  60

aac ctt ggg gag atg gtg gcc cag ctg agg aac agc tcc gag cca gcc   240
Asn Leu Gly Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu Pro Ala
 65                  70                  75                  80

aag aag aaa tgt gaa gtc aat cta cag ctg tgg ttg tcc aac aag agg   288
Lys Lys Lys Cys Glu Val Asn Leu Gln Leu Trp Leu Ser Asn Lys Arg
                 85                  90                  95

agc ctg tcc cca tgg ggc tac agc atc aac cac gac ccc agc cgc atc   336
Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser Arg Ile
            100                 105                 110
```

```
cct gcg gac ttg ccc gag gcg cgg tgc cta tgt ttg ggt tgc gtg aat   384
Pro Ala Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys Val Asn
        115                 120                 125

ccc ttc acc atg cag gag gac cgt agc atg gtg agc gtg cca gtg ttc   432
Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro Val Phe
    130                 135                 140

agc cag gtg ccg gtg cgc cgc cgc ctc tgt cct caa cct cct cgc cct   480
Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Gln Pro Pro Arg Pro
145                 150                 155                 160

ggg ccc tgc cgc cag cgt gtc gtc atg gag acc atc gct gtg ggt tgc   528
Gly Pro Cys Arg Gln Arg Val Val Met Glu Thr Ile Ala Val Gly Cys
                165                 170                 175

acc tgc atc ttc                                                  540
Thr Cys Ile Phe
            180
```

<210> 9
<211> 783
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(783)

```
<400> 9
atg ctg cag gca gag gct gtg caa agt ccc tgt ggc agg aag gag cat   48
Met Leu Gln Ala Glu Ala Val Gln Ser Pro Cys Gly Arg Lys Glu His
  1               5                  10                  15

gct gag tgt aag act gag aga aga cca gag gct gga aca cac aag aga   96
Ala Glu Cys Lys Thr Glu Arg Arg Pro Glu Ala Gly Thr His Lys Arg
             20                  25                  30

gag gaa aca ggg ggt aat ctg agg agg gag agc tgg ggg atg gta ggc   144
Glu Glu Thr Gly Gly Asn Leu Arg Arg Glu Ser Trp Gly Met Val Gly
         35                  40                  45

aga gtc aga cct gag cct ggg aga ccc cta aga gca atg gga aag gga   192
Arg Val Arg Pro Glu Pro Gly Arg Pro Leu Arg Ala Met Gly Lys Gly
     50                  55                  60

cca tgg acc cca agt ctt ggc cca gat aga gac ttg gct ctc cca cct   240
Pro Trp Thr Pro Ser Leu Gly Pro Asp Arg Asp Leu Ala Leu Pro Pro
 65                  70                  75                  80
```

```
ttg acc tgt ggc ttt cct ccc cag ctg ttt ctt ctt acc att tcc atc   288
Leu Thr Cys Gly Phe Pro Pro Gln Leu Phe Leu Leu Thr Ile Ser Ile
                85                  90                  95

ttc ctg ggg ctg ggc cag ccc agg agc ccc aaa agc aag agg aag ggg   336
Phe Leu Gly Leu Gly Gln Pro Arg Ser Pro Lys Ser Lys Arg Lys Gly
            100                 105                 110

caa ggg cgg cct ggg ccc ctg gcc cct ggc cct cac cag gtg cca ctg   384
Gln Gly Arg Pro Gly Pro Leu Ala Pro Gly Pro His Gln Val Pro Leu
        115                 120                 125

gac ctg gtg tca cgg atg aaa ccg tat gcc cgc atg gag gag tat gag   432
Asp Leu Val Ser Arg Met Lys Pro Tyr Ala Arg Met Glu Glu Tyr Glu
    130                 135                 140

agg aac atc gag gag atg gtg gcc cag ctg agg aac agc tca gag ctg   480
Arg Asn Ile Glu Glu Met Val Ala Gln Leu Arg Asn Ser Ser Glu Leu
145                 150                 155                 160

gcc cag aga aag tgt gag gtc aac ttg cag ctg tgg atg tcc aac aag   528
Ala Gln Arg Lys Cys Glu Val Asn Leu Gln Leu Trp Met Ser Asn Lys
                165                 170                 175

agg agc ctg tct ccc tgg ggc tac agc atc aac cac gac ccc agc cgt   576
Arg Ser Leu Ser Pro Trp Gly Tyr Ser Ile Asn His Asp Pro Ser Arg
            180                 185                 190

atc ccc gtg gac ctg ccg gag gca cgg tgc ctg tgt ctg ggc tgt gtg   624
Ile Pro Val Asp Leu Pro Glu Ala Arg Cys Leu Cys Leu Gly Cys Val
        195                 200                 205

aac ccc ttc acc atg cag gag gac cgc agc atg gtg agc gtg ccg gtg   672
Asn Pro Phe Thr Met Gln Glu Asp Arg Ser Met Val Ser Val Pro Val
    210                 215                 220

ttc agc cag gtt cct gtg cgc cgc cgc ctc tgc ccg cca ccg ccc cgc   720
Phe Ser Gln Val Pro Val Arg Arg Arg Leu Cys Pro Pro Pro Pro Arg
225                 230                 235                 240

aca ggg cct tgc cgc cag cgc gca gtc atg gag acc atc gct gtg ggc   768
Thr Gly Pro Cys Arg Gln Arg Ala Val Met Glu Thr Ile Ala Val Gly
                245                 250                 255

tgc acc tgc atc ttc                                             783
Cys Thr Cys Ile Phe
            260
```

<210> 10
<211> 21
<212> DNA
<213> synthetic construct

<400> 10
ttaccatttc catcttcctg g 21

<210> 11
<211> 18
<212> DNA
<213> synthetic construct

<400> 11
tgcaggtgca gcccacag 18

<210> 12
<211> 21
<212> DNA
<213> synthetic construct

<400> 12
gcatggtgaa gggattcacg c 21

<210> 13
<211> 265
<212> DNA
<213> synthetic construct

<400> 13
gaatggaaga gtatgagcgg aaccttgggg agatggtggc ccagctgagg aacagctccg 60

agccagccaa gaagaaatgt gaagtcaatc tacagctgtg gttgtccaac aagaggagcc 120

tgtccccatg gggctacagc atcaaccacg accccagccg catccctgcg gacttgcccg 180

aggcgcggtg cctatgtttg ggttgcgtga atcccttcac catgcaggag gaccgtagca 240

tggtgagcgt gccagtgttc agcca 265

<210> 14
<211> 155
<212> PRT
<213> Homo sapiens

<400> 14
Met Thr Pro Gly Lys Thr Ser Leu Val Ser Leu Leu Leu Leu Leu Ser
1 5 10 15

Leu Glu Ala Ile Val Lys Ala Gly Ile Thr Ile Pro Arg Asn Pro Gly
20 25 30

Cys Pro Asn Ser Glu Asp Lys Asn Phe Pro Arg Thr Val Met Val Asn
35 40 45

```
Leu Asn Ile His Asn Arg Asn Thr Asn Thr Asn Pro Lys Arg Ser Ser
     50                  55                  60

Asp Tyr Tyr Asn Arg Ser Thr Ser Pro Trp Asn Leu His Arg Asn Glu
 65                  70                  75                  80

Asp Pro Glu Arg Tyr Pro Ser Val Ile Trp Glu Ala Lys Cys Arg His
                85                  90                  95

Leu Gly Cys Ile Asn Ala Asp Gly Asn Val Asp Tyr His Met Asn Ser
            100                 105                 110

Val Pro Ile Gln Gln Glu Ile Leu Val Leu Arg Arg Glu Pro Pro His
        115                 120                 125

Cys Pro Asn Ser Phe Arg Leu Glu Lys Ile Leu Val Ser Val Gly Cys
    130                 135                 140

Thr Cys Val Thr Pro Ile Val His His Val Ala
145                 150                 155


<210> 15
<211> 158
<212> PRT
<213> Mus musculus

<400> 15
Met Ser Pro Gly Arg Ala Ser Ser Val Ser Leu Met Leu Leu Leu Leu
  1               5                  10                  15

Leu Ser Leu Ala Ala Thr Val Lys Ala Ala Ala Ile Ile Pro Gln Ser
             20                  25                  30

Ser Ala Cys Pro Asn Thr Glu Ala Lys Asp Phe Leu Gln Asn Val Lys
         35                  40                  45

Val Asn Leu Lys Val Phe Asn Ser Leu Gly Ala Lys Val Ser Ser Arg
     50                  55                  60

Arg Pro Ser Asp Tyr Leu Asn Arg Ser Thr Ser Pro Trp Thr Leu His
 65                  70                  75                  80

Arg Asn Glu Asp Pro Asp Arg Tyr Pro Ser Val Ile Trp Glu Ala Gln
                85                  90                  95

Cys Arg His Gln Arg Cys Val Asn Ala Glu Gly Lys Leu Asp His His
            100                 105                 110

Met Asn Ser Val Leu Ile Gln Gln Glu Ile Leu Val Leu Lys Arg Glu
        115                 120                 125

Pro Glu Ser Cys Pro Phe Thr Phe Arg Val Glu Lys Met Leu Val Gly
    130                 135                 140
```

Val Gly Cys Thr Cys Val Ala Ser Ile Val Arg Gln Ala Ala
145 150 155

<210> 16
<211> 151
<212> PRT
<213> Herpesvirus saimiri

<400> 16
Met Thr Phe Arg Met Thr Ser Leu Val Leu Leu Leu Leu Leu Ser Ile
1 5 10 15

Asp Cys Ile Val Lys Ser Glu Ile Thr Ser Ala Gln Thr Pro Arg Cys
20 25 30

Leu Ala Ala Asn Asn Ser Phe Pro Arg Ser Val Met Val Thr Leu Ser
35 40 45

Ile Arg Asn Trp Asn Thr Ser Ser Lys Arg Ala Ser Asp Tyr Tyr Asn
50 55 60

Arg Ser Thr Ser Pro Trp Thr Leu His Arg Asn Glu Asp Gln Asp Arg
65 70 75 80

Tyr Pro Ser Val Ile Trp Glu Ala Lys Cys Arg Tyr Leu Gly Cys Val
85 90 95

Asn Ala Asp Gly Asn Val Asp Tyr His Met Asn Ser Val Pro Ile Gln
100 105 110

Gln Glu Ile Leu Val Val Arg Lys Gly His Gln Pro Cys Pro Asn Ser
115 120 125

Phe Arg Leu Glu Lys Met Leu Val Thr Val Gly Cys Thr Cys Val Thr
130 135 140

Pro Ile Val His Asn Val Asp
145 150

<210> 17
<211> 18
<212> DNA
<213> synthetic construct

<400> 17
gcccagctga ggaacagt 18

<210> 18
<211> 18
<212> DNA

<213> synthetic construct

<400> 18
gagctggccc agagaaag 18

<210> 19
<211> 18
<212> DNA
<213> synthetic construct

<400> 19
acaggaacct ggctgaac 18

<210> 20
<211> 18
<212> DNA
<213> synthetic construct

<400> 20
agcccacagc gatggtct 18

<210> 21
<211> 8
<212> PRT
<213> synthetic construct

<400> 21
Asp Tyr Lys Asp Asp Asp Asp Lys
1 5

<210> 22
<211> 33
<212> DNA
<213> synthetic construct

<400> 22
aaaaaagaat tcgccgccat ggactggccg cac 33

<210> 23
<211> 54
<212> DNA
<213> synthetic construct

<400> 23
tcctctgtcg actcacttgt cgtcgtcgtc cttgtagtcg aagatgcagg tgca 54

<210> 24
<211> 31

<212> DNA
<213> synthetic construct

<400> 24
tttttagatc tcggaacacc aaaggcaaaa g 31

<210> 25
<211> 30
<212> DNA
<213> synthetic construct

<400> 25
ttttgtcgac tcagaagatg caggtgcaac 30

<210> 26
<211> 31
<212> DNA
<213> synthetic construct

<400> 26
gatcgatcga tcccatggac tggcctcaca a 31

<210> 27
<211> 65
<212> DNA
<213> synthetic construct

<400> 27
tgactgactg actgtcgact cactacttgt catcgtcatc cttatagtcg aagatgcagg 60

tgcag 65

<210> 28
<211> 24
<212> DNA
<213> synthetic construct

<400> 28
gactataagg atgacgatga caag 24

<210> 29
<211> 10
<212> PRT
<213> synthetic construct

<400> 29
Arg Ser Pro Lys Ser Lys Arg Lys Gly Gln
1 5 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/27068

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :536/23.5, 387.1; 435/252.3, 325, 69.1, 7.1, 6; 530/351
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.5, 387.1; 435/252.3, 325, 69.1, 7.1, 6; 530/351

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

1993/1994 Promega Catalog

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MPSRCH, APS, STN, SEQ ID NOs 1-6, il17, il(w)17, interleukin(w)17, interleukin17, embryo(w)derived(w)interleukin#, edirf

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X - Y	Database Genbank, Accession No. AA044549, MARRA, M. et al. The WashU- HHMI Mouse EST Project. 05 September 1996, see entire document.	1, 3, 5 -------- 16-18
Y	Promega 1993/94 Catalog, Promega Corporation, 2800 Woods Hollow Road, Madison, Wisconsin, USA, 1993, page 153, see entire document.	16-18

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 24 FEBRUARY 1999	Date of mailing of the international search report 17 MAR 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer D. Lamer for JAMES MARTINELL Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/27068

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12N 15/24, 15/00, 15/85; A61K 38/20; C07K 16/00; C12P 21/02; G01N 33/53, 33/68; C12Q 1/68